The English Verb *Kill*, the Finnish Verb *Tappaa* and their Equivalents in the Other Language – a Corpus-based Contrastive Study

Mari Saarhelo
University of Tampere
School of Language, Translation and Literary Studies
English Philology
MA Thesis
April 2014

Tampereen yliopisto
Englantilainen filologia
Kieli- käännös- ja kirjallisuustieteen yksikkö

SAARHELO, MARI: The English Verb *Kill,* the Finnish Verb *Tappaa* and their Equivalents in the Other Language – a Corpus-based Contrastive Study

Pro Gradu -tutkielma, 73 sivua + liitteet [3 sivua]
Huhtikuu 2014

Tässä pro gradu -tutkielmassa tutkitaan englannin verbiä *kill*, suomen verbiä *tappaa* ja niiden vastineita tämän kieliparin sisällä. Tutkielman tarkoituksena on selvittää, millaisia vastineita *kill* ja *tappaa* saavat ja mitkä seikat vaikuttavat näiden vastineiden valintaan. Tätä varten tarkastellaan verbien erinäisiä syntaktisia ja semanttisia ominaisuuksia sekä tekstityyppejä, joissa ne esiintyvät. Tutkimus sijoittuu kontrastiivisen korpuslingvistiikan kenttään ja korpuspohjainen analyysi muodostaakin keskeisen osan sen sisällöstä.

Materiaaleina tutkimukselle olivat useat suomen- ja englanninkieliset sanakirjat sekä yksi korpus. Korpuksena käytettiin Tampereen yliopistolla kerättyä kaksikielistä käännöskorpusta *Tampere Bilingual Corpus of Finnish and English* (TamBiC). Kaksikielisten sanakirjojen pohjalta mukaan sanakirjamateriaalin tarkasteluun otettiin myös suomen verbi *surmata*. Kontrastiivinen analyysi tehtiin käyttäen Robert Ladon mallia. Sanakirjojen avulla tutkittiin verbien *kill*, *tappaa* ja *surmata* saamia merkityksiä ja niiden kieliopillisia ominaisuuksia. Jälkimmäisten määrittelyyn käytettiin Huddlestonin sanaluokkien määrittelyyn käytettävää jaottelua.

Sanakirjatutkimuksen pohjalta tehty kontrastiivinen analyysi osoitti, että yhdenkään tarkastellun verbin käyttöä eivät rajaa ainakaan täysin kirjaimellinen/kuvainnollinen merkitys, se, onko verbi transitiivinen vai intransitiivinen tai se, onko verbin objekti elollinen vai eloton. Tämän pohjalta siirryttiin korpusanalyysiin. Haut TamBiC:stä tuottivat 241 osumaa verbille *kill* ja 174 osumaa verbille *tappaa*. Näistä verbillä *tappaa* oli liian vähän vaihtelua englanninkielisissä vastineissaan, joten tarkempi korpusanalyysi keskittyi vain verbiin *kill*, jonka tärkeimmiksi vastineiksi löytyivät *tappaa* (152 osumaa), *surmata* (34 osumaa), *kuolla* (18 osumaa), rakenteet, joissa esiintyy sana *henki* (8 osumaa) ja *kaatua* (8 osumaa). Muita, harvinaisempia vastineita oli myös useita.

Korpusanalyysissä tarkasteltiin taivutusmuotoja, passiivia, semanttisia objekteja ja subjekteja, semanttisia merkityksiä ja tekstityyppejä. Tämän analyysin tulokset osoittautuivat lähes yhtä ei-sitoviksi kuin sanakirja-analyysinkin. Vastineista *kaatua* oli ainoa, jolle tulokset pystyivät osoittamaan tarkan käytön: kun *kill* on passiivissa ja tekstin aihealueena on sota. *Tappaa* oli yleisin vastine tarkasteltavasta ilmiöstä huolimatta käytännössä kaikissa tapauksissa. Lupaavin tutkituista ilmiöistä oli passiivi, jossa kaikkien muiden vastineiden suhteelliset osuudet nousivat ja *tappaa*:n laskivat. Näille ei-sitoville tuloksille on vaikeaa tarjota konkreettista käyttöä ja niiden voidaankin lähinnä katsoa osoittavan, että näiden nimenomaisten verbien kohdalla kielioppiin ja merkityksiin perustuva lähestymistapa ei ole sopivin. Ehdotettuja jatkotutkimussuuntia ovat ajallisen näkökulman ottaminen mukaan ja kyselytutkimus, jossa testataan halua käyttää eri vastineita erilaisissa tilanteissa.

Avainsanat: korpuslingvistiikka, kontrastiivinen lingvistiikka, käännösvastineet, verbit

Table of Contents

1. Introduction

The aim of this contrastive study is to find out what kind of criteria can be used to determine which Finnish equivalent is most suitable for the English verb *kill* and which English equivalent for the Finnish verb *tappaa* in different situations. This will be done mostly by examining the syntactical characteristics of *kill* and *tappaa* and their equivalents based on dictionaries and corpus data, but semantic qualities and text types will also be examined, for as noted by Baker (2011, 15), ”the choice of a suitable equivalent in a given context depends on a wide variety of factors”. She (2011, 6) also states that while the current trend in both translation and linguistics seems to to focus on the text as a whole, one needs to understand the lower level structures that make up the complete texts. Following this “bottom-up” principle, this study will start with the syntax and semantics in the dictionary analysis and then continue with those in the corpus analysis, which will then end in an analysis of text types.

The dictionary analysis will be done with the help of Huddleston's (1988) criteria for defining word classes, and end in a contrastive analysis based on Lado's (1957) model. Of the potential equivalents the study will examine *tappaa* and *surmata* for *kill* and *kill* itself for *tappaa* more closely, but in the corpus chapter, other equivalents found in the data will be taken into account as well. The corpus analysis is the main focus of this study, while the information given by the dictionaries is mostly used to set a basis for it. Thus, the main theoretical bakcgrounds for this thesis lie in corpus linguistics and contrastive linguistics.

Kill and *tappaa* were chosen because I have done my BA-thesis on *kill* and that study proved that more research needs to be done on the matter to achieve more conclusive results. In that study *kill* was chosen because preliminary searches showed that examining it would not end in a situation where only one equivalent would be found. *Tappaa* was added to the study to further examine the relationship between the two verbs and to find out whether it had similar variation in its equivalents as *kill* has.

1.1 Research questions and structure

I have three main research questions: 1. What Finnish equivalents does *kill* have? What English equivalents does *tappaa* have? 2.Which phenomenona affect the choice between these equivalents? 3. Are these phenomena syntactical, semantic or otherwise?

The structure of this study is the following: in this first chapter, I introduced the study in general and my research questions. There will also be an overview of previous research done into similar topics. Chapter two will give the theoretical background, discussing corpus linguistics and contrastive linguistics. In chapter three I will explain the methods that I will use to analyse my data. There I will also introduce the materials I gathered for the study: what they are, why these particular ones and how they were gathered. Chapter four will begin the actual analysis with examining what dictionaries have to say about *kill, tappaa* and *surmata,* whose inclusions I will explain in detail at the beginning of that chapter. Chapter five will present the contrastive analysis based on the dictionary findings. Chapter six will continue the analysis, but shift from dictionaries to corpus data. In chapter seven I will gather my main findings from the corpus analysis. Finally, in chapter eight I will conclude the study, discussing some possibilities for further research.

1.2 Previous studies

Angus McIntosh's ”A four-letter word in *Lady Chatterly's Lover*” (1960) is an early example of a contrastive study. If the definition of a corpus is not used very strictly, it can even be said to be corpus-based, as he uses authentic text, not intuition as his basis. (On the definition of corpus, see chapter 2.1). In his study, McIntosh examines the relationship between the English verb *know* and the French verbs *savoir* and *connaître*, using the novel *Lady Chatterley's Lover* as his data. He (1960, 151-152) was interested in finding out whether grammatical criteria recognisable by a computer could be used to make the choice between *savoir* and *connaître*. He focused on the types

of objects *know* had. Criteria that worked were *know* having: a clause or no object (1960, 152), specific types of pronoun objects (1960, 154), an animate or proper-name object (1960, 155) and finally ”inanimate indefinite[1]” object (1960, 158). These criteria in this order worked for all but two of his 293 instances of *know.* McIntosh (1960, 159) drew the conclusion that his study ”at least suggests the vital relevance of grammatical relationships in problems of this kind”.

In his book *Seeing Through Multilingual Corpora : On the Use of Corpora in Contrastive Studies* (2007), Stig Johansson also does a case study where he compares several nouns in Norwegian and English, mostly based on English-Norwegian Parallel Corpus of which he was one of the compilers. He has several research questions concerning the correspondence (a term he uses instead of equivalence) of the nouns between the languages, mostly about how they correspond and whether there are any notable differences such as lexical gaps or differences in word classes (Johansson 2007, 41). His conclusion is that when moving on to more complex (in that they are used in various ways and often idiomatically) nouns, correspondence tends to disappear more, especially when it comes to keeping the word class same. For example the English *mind* often has a verb as a Norwegian equivalent (Johansson 2007, 93). He also states (2007, 93) that his results show that “words have grammar”. The following chapters of Johansson's book are also various case studies where he examines *love* and *hate* (as verbs), spending time, *seem*, expressions of usuality, negation, generic personal pronoun *man*, replacement of subjects in translation, sentence openings, Norwegian concessive marker *likevel* and *well* (as a discourse particle). In most of these, the languages in contrast are English and Norwegian, but in some, German is also included. While Johansson's scope of studies is extensive, he states (2007, 301) that it “represents just a fraction of current work on multilingual corpora”, while introducing the work of several other researchers in the field, several of which go beyond contrastive linguistics.

Research of the type done in this study in Finland seems to be focused in the University of Tampere, as going through the Master's Thesis online databases of the country's other universities

1 Such as *nothing*, *all* or *anything*.

revealed little to no contrastive corpus-based research. On the other hand, master's theses similar to this one have been done in the University of Tampere. The first of the ones I found is Paula Suoniemi's thesis from 2006. As she examines words that are more grammatical than lexical, her point of focus is somewhat different from those of the previous studies. Suoniemi is interested in the concept of cohesion and how the discrepancy between the Finnish and English third person singular pronouns affects it and what strategies have been employed in her data to keep the cohesion. Suoniemi uses the same corpus I am using, (The Tampere Bilingual Corpus of Finnish and English, TamBiC, see chapter 3.2 for more information on the corpus), in a shorter form it was in 2006 and also encountered some difficulties in getting her data due to the fact that TamBiC was not in readily accessible form in 2006. Her findings were that both English and Finnish use the use of (proper) nouns and ellipsis as the main strategies for translation when the use of a pronoun would cause ambiguity or to avoid repetition.

In her thesis (2008), Saara Salminen examined the relationship between the English *environment, circumstances* and *surroundings*, and the Swedish *miljö*, *omständigheter* and *omgivning*. She focuses on the concept of synonymity and how it will affect the choosing of translation equivalents (chapter 2.1.1), but is also interested in collocation (chapter 2.1.2) Structurally her thesis is very similar to this one first doing dictionary analysis for finding the meanings and usages of the words (chapter 4.2), then corpus analysis (chapter 4.3). However, Salminen's focus does seem to be more on the synonymity inside the languages, dedicating much more space to it than actual contrastive study, and also including diachronic perspective. Her findings (2008, 78) are that in both languages, there are enough differences between the words that they cannot be said to be fully synonymous, making them near-synonyms instead. Further (2003, 78-79), the only somewhat clear equivalence was found between *circumstances* and *omständigheter*, and some fixed phrases.

Katariina Franssila's thesis (2010) is from the field of corpus linguistic translation studies,

not corpus-based contrastive linguistics, but does use a corpus when examining the English verb *manage* and its Finnish and German equivalents in the specific environment of the texts produced by the European Commission. She focuses on the collocation of *manage*, specifically the objects of it. Her findings were that *manage* has several different equivalents both in Finnish and German, but in German *verwalten* is the most preferred equivalent while in Finnish no one word rose to such popularity. *Verwalten* was used with different collocation groups while the four most popular Finnish equivalents, *hallinnoida, johtaa, hallita* and *hoitaa* differed somewhat in their collocations.

Closest to the current study is Pertovaara's thesis from 2012. In it she examines the verbs *work* and *play* and their Finnish equivalents and whether the different syntactical environments and collocations of the verbs could help distinguish between the meanings and thus help in choosing a particular equivalent. A specific aim is to find out whether this could be of use in machine translation. The structure of the study is very similar to this one. First a dictionary analyses of the different syntactic characteristics and meanings of the words examined followed by their Finnish equivalents in chapter 4, which end in contrastive analyses for both. Then there is a corpus analysis of *work* and *play* in chapter 5. This is based on TamBiC. However, differing from the structure of the current study, Pertovaara continues to organize the analysis based on the different meanings of the words, mostly because very distinctive meanings could be found for both *work* and *play.* She does include examination of inflections and dependents, however (chapters 5.3 and 5.6 of Pertovaara's thesis). Pertovaara's conclusion (chapter 6) is that for both *work* and *play*, distinctive syntactic and collocational environments could be found for the different meanings, ones that could potentially be of use to machine translation.

2. Theoretical background

Lewandowska-Tomaszczyk and Dziwirek (2010, 5) express nicely the usefulness of combining contrastive and corpus linguistics: ”In the contrastive perspective, what is achieved is cross-

linguistic contextual analysis of language pairs. Large and varied corpus data lend support to positing a cline of narrowly and broadly conceived cross-language equivalents".

2.1 Corpus linguistics

To be able to discuss what corpus linguistics is, one must first define what a corpus is. According to Charles Meyer (2002, xi) a linguistic corpus can be defined as "a collection of texts or parts of texts upon which some general linguistic analysis can be conducted". Based on this, it might seem like the corpus is just a useful tool for various types of linguistic research and corpus linguistics just a methodology. According to Laviosa (2002, 8-9) this is not the case, as corpus linguistics does include its own theoretical statements about the nature of language and how it should be researched, namely: the use of authentic data, focus on empirical study, and the idea that language is a live thing, ever changing, its variations just as important and interesting as its more typical uses.

As Svartvik's (2007) article on the history of corpus linguistics demonstrates, this field of linguistics was not always popular, as introspection was the most used method when corpora first began to be used. In 1992 (137) Leech and Fligelstone expressed the belief that computer corpora will become increasingly more used in linguistic research. This has happened. Svartvik (2007, 12) states that corpus linguistics has changed from "fringe activity" to "mainstream methodology". Paul Baker (2009, 1) describes how the use of corpora has also been increasing in other areas of linguistics and even beyond. Also noteworthy is the fact that, as Svartvik (2007, 15) points out, a non-native speaker of a language can find it impossible to rely on introspection for study. This study examines two languages, which in my opinion makes the use of corpora necessary, as truly bilingual users of English and Finnish are quite rare.

As the corpus used in this study is a bilingual one, this specific type of corpus needs its own definition. The one offered by Johansson (2007, 9) is "a collection of texts in two or more languages put together in a principled way" "parallel in some sense, either by being in a translation

relationship or by being matched with respect to genre, time of publication, degree of formality, etc.”. In the case of the corpus used in this study (TamBiC), the parallelism comes from the texts being translations. (TamBiC will be discussed in more detail in chapter 3.2). The specific name Johansson offers for this type of corpora is ”translation corpora”. The other type of parallel corpus is the comparable corpus, where texts have been matched by other similarities such as genre or time of publication (Johansson 2007, 9-10). There has been some variation in the use of the terms “parallel corpus”, “comparable corpus” and “translation corpus” between corpus linguists and contrastive linguists, with “parallel corpus” being especially problematic, as it has been used to refer to either of the previously mentioned types or both of them (Laviosa 2002, 37). The approach used in this study is a combination, where “parallel corpus” is the umbrella term, under which “translation corpus” (texts and their translations) and “comparable corpus” (texts matched in other ways besides translation) can be found.

Johansson (2007, 9) gives a short overview of the good and bad sides of using a translation corpus, the main problem being the fact that a translation is always a result of translators handling texts according to their own preferences and ideas. This means that the choices of an individual can greatly affect the resulting text, and thus the corpus data. In comparable corpora the main problem is defining what is truly comparable (Johansson 2007, 9). However, TamBiC includes texts both originally Finnish and English and their translations (again, see Chapter 3.2). This is a structure similar to Johansson's and Hofland's English-Norwegian Parallel Corpus (ENCP). According to Johansson (2007, 11-12) this way of making a parallel corpus helps overcome the problems of both types, as translated language can be compared to the texts originally in that language to find out which features are results of translation, while the texts remain clearly comparable. The largest problem with this type of corpus is one that TamBiC suffers from as well: small size caused by the lack of suitable texts, and also somewhat limited text types. Especially when one of the languages in a bidirectional translation corpus is much more translated out of than the other, as is the case of

almost every corpus where one of the languages is English, finding suitable texts for both halves of the corpus is difficult. ENCP encountered this problem as well as did CEXI, an Italian-English corpus (Laviosa 2002, 41).

2.2 Contrastive linguistics

Johansson (2007, 1) gives contrastive analysis the definition ”systematic comparison of two or more languages, with the aim of describing their similarities and differences”. Krzeszowski (1990, 10) adds to this that these languages need not be related. Granger (2003, 17) describes the history of contrastive linguistics as ”a pattern of success-decline-success”. This is because it was at first used mainly for foreign language learning, and after a while it was realized that contrastive analysis does not work for that purpose as well as was thought (ibid.). Johansson (2007, 2) gives some more details: the idea was that specific language pairs would cause specific problems in learning because of their differences and similarities. Recognising these would help in the learning, as more effort could be focused on those areas deemed different and thus difficult. This error analysis is what Lado focused on in his *Linguistics Across Cultures*, the subtitle of which (*Applied linguistics for language teachers*) is telling. According to Krzeszowski (1990, 189), Lado hoped that contrastive analysis would help in error analysis because it would help in recognising the places where positive or negative transfer would happen. Later on it was realized that ”language learning cannot be understood by a purely linguistic study” (Johansson 2007, 2) and that there were more difference between error analysis and contrastive analysis than Lado thought (Krzeszowski 1990, 190). There was also the problem that this earlier incarnation of contrastive linguistics (or contrastive analysis, as it was then called), was connected to introspection, structuralism and generative grammar and while they fell out of favour, so did contrastive linguistics (Altenberg and Granger 2002, 6). In 1990, Krzeszowski (1) begins his book on contrastive linguistics by stating that it does not “enjoy much respect among linguists”. One might ask whether Lado, as a prime example of this earlier line

of contrastive analysis, can be used as a basis for a modern contrastive study. But while his theory might be outdated, in my opinion his method is not. This model will be discussed in more detail in chapter 3.3. And Lado himself (1957, 73) says that while the main applications for the analyses done by comparing languages are directly in teaching, they can be “written up for publication or other distribution”, implying that he did not think other uses impossible.

Granger (2003, 18) states that two factors helped the revival of contrastive linguistics: applying it to fields other than education (such as machine translation or lexicography), and the increasing use of corpora. The latter of these gave ”a much more solid empirical basis” to the research, as opposed to the earlier, intuition-based, model (ibid.). Gonzales et al. (2008, XVI) refer to the enhanced ”testability, authenticity and general empirical adequacy” of corpora use. In an another book (2002, 6) Altenberg and Granger offer one further explanation for the new interest in contrastive linguistics: ”increasing demand for multilingual and cross-cultural competence” especially in Europe. Barlow (2008, 101) actually states that this European multilingualism meant that contrastive studies remained if not popular, at least used here when their use declined in America. According to Altenberg and Granger (2002, 6) in the seventies, when contrastive analysis had already fallen out of favour in the USA, there were large projects comparing English to other European languages in Europe. This study uses all three of these factors: its aim is not related to education and language acquisition, it uses a corpus study as its main component, and the languages it contrasts are two European ones, Finnish and English.

2.3 On equivalence, translations and synonymy

It should be noted that the term ”equivalent”, frequently used throughout this study, is not entirely uncontroversial. First of all, it is a term fundamentally based in translation studies and the full form given could actually be ”translation equivalent”. Baker's *In Other Words: A Coursebook on Translation* (2011) makes equivalence the main topic of translation, having chapters such as

”Equivalence at word level” or ”Grammatical equivalence”. This being a linguistic study instead of translation one raises the question whether the term should be used at all. It is however, as Lewandowska-Tomaszczyk and Dziwirek (2010, 6) put it ”of fundamental importance for contrastive analysis”. Secondly, even in translation studies, not every researcher prefers to use the term. But in Kenny's (2009) overview of the term, the problems she brings up are actually ones that work for this study instead of against it. According to her (2009, 96) some translation researchers dislike the term because of its connection to linguistic approaches to translation and because it is somewhat circular: translation and equivalence define each other. The present study, as discussed earlier, is linguistic in orientation and when the corpus data is analysed later on in chapter 6, the direction of the translation is not discussed, in short is not deemed important which of the two languages is the source and which the target language. This later choice was made because the aim of this study is not to discuss the properties of translated language per se and also because of the small size of the dataset (more on this in chapter 3.3). This choice is also the reason the term ”translation” is avoided in the study. The equivalents of *kill* and *tappaa* are not translations in several of the cases they are referred to (in these cases *kill* and *tappaa* are the translations), and calling them translations in these cases would be misleading. Equivalence works both ways, and makes for a consistent term for the whole study because of this.

According to Krzeszowski (1990, 16), in studies of lexical (or syntactic) level, semantic equivalence can be used as *tertium comparationis*, the common point of reference for the two things compared. However Krzeszowski also notes that translation equivalence and semantic equivalence are actually two different things, arguing that previous authors are actually mistakenly using the two interchangeably. Thus, his term ”pragmatic equivalence” (1990, 18) is closer to how the term is used in this study, and avoids going into too intricate analysis of terms that are mostly complicated when applied to larger units than single words or short phrases, as is the case in this study. Baker (2011, 230) in her chapter on pragmatic equivalence defines pragmatics as meaning, but from the

participants' perspective, not something inherent in the text. Thus, words or phrases have pragmatic equivalence if the reader can get from them the same meaning in both languages in a particular context. The term ”textual equivalence” is also quite fitting for the use of the term in this study. This term was first used by Catford in 1965 (Kenny, 2009, 98), and is thus somewhat old-fashioned, but Kenny also states (2009, 99) that it is ”the basis of much contemporary work in contrastive linguistics”, among others. It is an empirical form of equivalence, where the texts in both languages have been already produced and their equivalence is determined by direct comparison by bilingual informants (Kenny, 2009, 98-99). On surface level it looks like pragmatic equivalence and textual equivalence are contradictory terms. “Textual equivalence” as a term seems to restrict equivalence to the texts themselves without leaving room for context or reader interpretation, but this is not actually the case. Kenny (2009, 99) places the term into the larger framework of equivalence as an empirical phenomenon, descriptive translation studies, where Toury (1995, 61) argues that equivalence need to be determined while taking its circumstances (i.e. context) into account.

A final note should be made on the relationship between the concepts of synonymy and equivalence. On the surface, the two seem very similar, as they both deal with similarity. But, first of all, equivalence can be concerned about things much larger than the word- or phrase-level. This already be seen by looking at the names of the chapters in *In Other Words* (Baker 2011). While chapter 2 is ”Equivalence at world level”, she discusses things such as collocations, grammatical categories and cohesion in the following chapters, while still still using the term “equivalence” throughout all of this. In Kenny's (2009, 97) overview of the subject she moves from researchers referring to source and target text words as equivalent to those referring to the texts themselves as wholes as equivalents. Synonymy on the other hand is a fundamentally lexical concept (Cruse 1986, 267). While the different equivalents of a word are often near-synonyms, and can be examined from this perspective, as was done in Salminen's thesis (chapter 1.2), this is not the approach taken in this study. A different approach was chosen mostly to not restrict the examination

to those equivalents that are immediately apparent and listed in thesauri. After all, one of the research questions is finding out what equivalents have actually been used for *kill* and *tappaa*.

3. Materials and methods

3.1 Dictionaries

Three English dictionaries were used in this study. They are *Collins COBUILD Advanced Learner's Dictionary* (referred from now on as COBUILD), *Longman Dictionary of the English Language* (LDEL) and *Oxford Advanced Learner's Dictionary of Current English* (OALD). Some larger dictionaries were not chosen because this study is focused on the most common senses of *kill*, as those are more likely to appear in the corpus data examined, and the dictionaries chosen cover those meanings well enough. Likewise, there is no diachronic aspect to the study, and these two factors combined made the use of more comprehensive dictionaries like *The Oxford English Dictionary unnecessary.*

The number of existing Finnish dictionaries is much smaller than English ones, so there was not much room for choice among them, but three useful ones were found as well: *Kielitoimiston sanakirja* (KTS), *Nykysuomen sanakirja* (NSS) and *Suuri suomen kielen sanakirja* (SSKS).

Dictionaries were used to find what characteristics *kill, tappaa* and their equivalents have according to them. This can be seen as a preliminary study to the corpus one, which is the main focus of this thesis. The information given in dictionaries can be used to guide the corpus study to particular areas by, for example, noting a clear difference in the ways some equivalents are used. It is also interesting to see whether there are any differences between the dictionary information and the results of the corpus study.

3.2 Corpus

This is a lexical study, and a computer corpus is a particularly fitting base for it, for as Altenberg

and Granger (2002, 4) state “lexis lends itself perfectly to the form-based research at which computers excel”. Using terminology from Altenberg and Granger (2002, 7-8), TamBiC (short for The Tampere Bilingual Corpus of Finnish and English) is a bilingual, bidirectional translation corpus. Bilingual in this context means that TamBic is only concerned with two languages, instead of multiple ones. Bidirectional translation corpus means that it includes both texts originally in Finnish and texts originally in English, both with their translations into the other language. TamBiC consists of two million words There is no spoken section, but the source texts are of various types, both fiction and non-fiction, the latter of which include both books and news texts. TamBiC has been aligned on the sentence level, but has not been parsed. It can be searched with text strings, wild card symbol * and Boolean operators AND, OR, NOT and THEN. Both subcorpora (Finnish-English and English-Finnish) can be searched separately, and refined searches within results can be made. (https://www12.uta.fi/tambic/user_manual.html)

The corpus is somewhat small by modern standards, but unfortunately a larger one handling this particular pair of languages does not exist. Finnish and English are not a very common pair of languages to study, and because of this there were not many corpora to choose from. TamBiC was chosen because of the following reasons: it includes a wide variety of text types (as opposed to some more specific corpora, such as European Parliament Proceedings Parallel Corpus (http://www.statmt.org/europarl/)), which allows me to examine text type as one of the criteria for choosing an equivalent. I have also done some previous work with it, which made using it easier and faster. *Kill* and *tappaa* are not the most common words of their languages, and the number of results received from corpus searches for them in TamBiC was not very high. Because of this it can be asked whether the results of this study are actually usable and valid, and can any conclusions about the equivalents of *kill* and *tappaa* be drawn based on them. My view on the matter is that while the results might not be as conclusive as those based on more data, they can at least be used as guidelines for further study. In itself, it might be interesting to see how much the results would

differ if the corpus used was changed. And while the verbs are not the most common ones, they are used in standard language, and are not limited by formality or other factors like it.

3.3 Methods

The contrastive analysis this thesis will follow Lado's model. According to him (1957, 66), there are three general categories to be taken into notice: form, meaning and distribution. Distribution will not be discussed, as it refers to which words the structure can be used with, and here the structure examined *is* a word. Lado's specific instructions (1957, 67-69) for natural languages are the following: 1. ”Locate the best structural description of the languages involved”, 2. ”Summarize in compact outline form all the structures” and 3. ”Actual comparison of the two language structures”. In this study, the first of these will be done with the help of dictionaries for meaning and Huddleston's grammar model of verbs for form. The second instruction will be used in a somewhat shortened form, as Lado seems to focus on sentence level, which allows for much more complicated structures to be relevant than the word level this study focuses on. The third instruction will be followed with the help of a table, collecting all the applicable information found in the dictionaries for overview. The form of the table is modified from Lado (1957, 73), who suggests two columns, one with the structures of one language and the other with corresponding structures of the second language on the same rows. Here the comparison will be done by giving a structure *kill* can occur in with a short example sentence, and then noting which of the two equivalents, *tappaa* and *surmata* can also occur in an equivalent Finnish structure.

Searching for instances of *kill* from the corpus was quite straightforward, and was done with the search string ”kill* NOT killer*”, where the asterisks allow for inflection. After this, the data was gone through manually to exclude instances that do not fit into the scope of the study, namely those where *kill* was a noun and those where the sentence it was in was not translated at all or translated too loosely to allow for analysis. This was necessary as TamBiC is not parsed in any way.

After this pruning, 244 sentence pairs were left, of which 145 were from originally English and 99 from originally Finnish texts. These two parts will mostly not be separated in the statistics or analysis, except when some specific sentence pair seems particularly interesting to examine from this point of view or the translation choices can not be explained without referring to which language was the original.

Due to the much more complicated system of conjugation of Finnish, ensuring that all the instances of all the conjugated forms of *tappaa* were found was more demanding than for *kill*. Simply searching with the root ”tap*” would not have worked due to the amount of various other words beginning with those letters, such as *tapahtua* or *tapaus*. The solution found was to make multiple searches for the various forms of *tappaa*, all conjugated partially to some extent. This reduced the amount of other words in the data, as the cut-off points were chosen so that as little of them as possible would get in, but it also made the amount of searches manageable as opposed to searching every possible conjugation on its own. Every search string ended with the ”wild card” marker ”*” to catch cases where multiple suffixes were used, such as *tappaisinko*, which would have been found with the search string ”tappa*”. The searches, after a pruning process similar to *kill*, resulted in 174 instances of *tappaa*, 104 originally English, 70 originally Finnish.

4. Dictionary analysis

Surmata was chosen to be examined here as well as *kill* and *tappaa* because according to *Englanti-Suomi Suursanakirja tappaa* and *surmata* are the most basic equivalents of *kill* giving them at the beginning of the entry for *kill* as a verb (s.v. *kill* 1), without giving any indication on how to choose between them. While other equivalents of *tappaa* besides *kill* are mentioned in the *Suomi-Englanti* Suursanakirja (s.v. *tappaa*), they are not treated as equals to it, and because of this will not be examined here.

4.1. Meanings

In this chapter I will examine what meanings the dictionaries discussed in chapter 3.1 give to *kill*, *tappaa* and *surmata*. A comparison of the meanings will follow in chapter 5 where I begin the contrastive analysis.

4.1.1. *Kill*

Oxford Advanced Learners Dictionary (OALD) gives ”to make sb/sth die” as the first meaning of *kill*, and provides this example among others:

> He tried to kill himself with sleeping pills. (OALD s.v. *kill* 1)

The second main meaning is ”to destroy or spoil sth or make it stop”:

> Do you agree that television kills conversation? (OALD s.v. *kill* 2)

Moreover, *kill* has some other, mostly informal or specialist meanings. For example, in informal language, *My feet are killing me*, according to the OALD, means ”My feet are causing me great pain”. Some dictionaries, however, like *COBUILD*, separate a semi-figurative meaning (”emphasizing that you are extremely angry with someone”) from the first main meaning, although the OALD gives this example under the first meaning of *kill*:

> My mother will kill me when she finds out.

An example of the fixed phrases used with *kill* in them would be *if it kills me*, which means ”emphasizing that you are determined to do [something] even though it is extremely difficult or painful” (COBUILD s.v. *kill* 10).

4.1.2. *Tappaa*

KTS states that the main meaning of *tappaa* is ”aiheuttaa jkn t. jnk kuolema...” [cause someone's or something's death] and adds a list of synonymous expressions, for example ”ottaa hengiltä, …, murhata, teurastaa”. The following is an example of this meaning:

> Kissa tappoi hiiren. [A cat killed a mouse.] (KTS)

Figuratively, *tappaa* means ”lopettaa, tukahduttaa, hävittää” and ”uuvuttaa, näännyttää”, exemplified in the following:

> Ankara kuri tappoi lasten aloitekyvyn. [Strict discipline killed the children's initiative.] (KTS)

4.1.3 S*urmata*

Surmata is defined in KTS as ”ottaa hengiltä, tappaa, ..., aiheuttaa kuolema” [take a life, kill, …, cause death]:

> Maanjäristys surmasi tuhat ihmistä. [The earthquake killed a thousand people.]

No figurative use is mentioned in KTS, but two other Finnish dictionaries NSS and SSKS give the following examples of the figurative use of *surmata*:

> Surmata joku katseellaan. [Kill someone with your look.] (SSKS)
>
> Surmasi haluttomuudellaan vanhempiensa toiveet. [His unwillingness killed his parents' hopes.] (NSS)

Interestingly, NSS also provides some information that would help in deciding between *tappaa* and *surmata*: *surmata* is described as having a meaning that is ”vähemmän karkea” [less coarse] than that of *tappaa*. This is an aspect that will be further examined in the corpus analysis chapter below.

4.2. Syntactic characteristics

According to Huddleston (1988, 27), the properties of word classes can be categorized under four separate headings: function, dependents, inflection and lexical morphology. Because there is nothing distinctive about the form of the word *kill* that marks it as especially verb-like, such as an -ise-suffix, the fourth property is not useful for the purposes of this study and will not be discussed further. The first three will be used to define the characteristics of *kill* as a verb.

4.2.1. *Kill*

a) Functions

According to Huddleston (1988, 28) the most common function of verbs is to head kernel clauses. *Kill* does indeed have this function, as the dictionary examples reveal:

> The defeat last night killed the team's chances of qualifying. (OALD s.v. k*ill* 2)

Huddleston then goes on to state that verbs also head many other kinds of non-kernel clauses. Of the ones he mentions, *kill* heading a clause ”functioning as the complement of a modal operator” and ”a clause marked by the infinitival particle *to*” can easily be found in the dictionaries:

> Heroin can kill. (COBUILD s.v. *kill* 1)
> I bought a spray to kill the weeds. (OALD s.v. *kill* 1)

While the dictionaries do not give any examples of *kill* used in an imperative sentence, it is not too difficult to find it in this function:

> Kill her now! (BNCWeb, FRD 620)

b) Inflections

Verbs in English have six different inflectional forms (Huddleston 1988, 27). Here are examples

with *kill*, taken from the dictionaries and the BNC:

1.Base form:
*Don't **kill** yourself trying to get the work done by tomorrow.*
(OALD s.v. *kill* 1)

2.General present tense:
*Shee-it, I **kill** myself sometimes*.
(BNCWeb)

3.Present tense 3rd pers. sing:
*Cancer **kills** thousands of people every year.*
(OALD s.v. *kill* 1)

4.Past tense:
*The earthquake **killed** 62 people*.
(COBUILD s.v *kill* 1)

5.Present participle:
*He would soon launch the second offensive, **killing** off the peace progress*.
(COBUILD s.v. *kill* 3)

6.Past participle:
*More than 1,000 people have been **killed** by the armed forces*. (COBUILD s.v. *kill* 1)

Kill is a regular verb, i.e. it has no irregular inflections, and it is found in all the possible inflected forms.

c)Dependents

”Verbs take a wide range of dependents” (Huddleston 1988, 28), perhaps because they head a varying range of clauses. As a verb-specific dependent, Huddleston (1988, 29) mentions objects (with transitive verbs). This specificity makes objects the most interesting and suitable dependents for the purposes of this study. *Kill* can in fact be used both intransitively (without an object) and transitively:

> Tiredness while driving can kill. (OALD s.v. *kill* 1)
>
> His objective was to kill the space station project altogether. (COBUILD s.v. *kill* 3)

When *kill* has an object, that object can be either animate or inanimate. LDEL states (s.v. *kill*-synonyms) that literal and figurative *kill* take different types of objects, literal taking ”people, creatures and plants” and figurative ”plans, feelings and similar things”.

4.2.2. *Tappaa*

None of the dictionaries provide much explicit information about the syntactic characteristics of *tappaa.* However, some other characteristics can be inferred from the examples, as in the case of *kill*. Nevertheless, examples of the different functions are not easy to find, as the Finnish dictionaries give relatively few whole sentences as examples.

a) Functions

It is quite easy to find examples of *tappaa* heading a kernel clause in the Finnish dictionaries:

> Kissa tappoi linnunpoikaset pesäänsä. [The cat killed the chicks in their nest.] (NSS)
>
> Tapoin kalan vääntämällä niskat nurin. [I killed the fish by breaking its neck.] (SSKS)

Tappaa functioning as the complement of a modal auxiliary can also be found, though this use is much rarer in the dictionaries:

> Tauti ei kyennyt tappamaan hänen tarmoaan. [The disease could not kill her energy.] (NSS)

b)Inflections

Several tensed/inflected forms are found with *tappaa*. As Finnish has many more inflections than English, a list of all of them would be too long to give here, but here are some samples:

> Huonot elokuvat ja romaanit *tappavat* halun harrastaa todellista taidetta. [Bad movies and novels kill the desire to take interest in true art.] (NSS)

Tappoi miehen puukolla. [Killed a man with a knife.] (KTS)

Ankara verotus *on tappanut* yrityksiä. [Severe taxation has killed companies.] (KTS)

c) Dependents

Tappaa, like *kill*, can also be used both intransitively and transitively. A classic example of the former use would be *Älä tapa*, which is also given as an example in NSS. This is not a very prominent use, however, and there are very few examples of it in the dictionaries.

Tappaa takes a wide range of objects, though usually the literal sense is reserved for animate objects:

Pyörremyrsky tappoi kymmeniä *ihmisiä*. [The tornado killed tens of people.] (KTS)

Karhu tapettiin pesäänsä. [The bear was killed in its den.] (NSS)

Figurative *tappaa* can also take animate objects like literal *tappaa*:

Pitkä ylämäki tappoi juoksijat. [The long uphill killed the runners.] (NSS)

It can also take inanimate objects, for example *innostus* (SSKS) and *yritys* (KTS) KTS also has one example where the object of *tappaa* is not clearly animate, even though the example is under the literal use:

Puustoa tappavat saasteet. [Pollution kills the trees.]

The animacy of plant life is arguable, however, and thus this is hardly a clear-cut example of an inanimate object with *tappaa* in its literal sense.

4.2.3. *Surmata*

Syntactic information about *surmata* is as scarce as it is with *tappaa.* The only feature mentioned is in NSS, where it is stated that *surmata* occurs ”tav. henkilöobj:n ohella” [usually with a person object], at least (apparently) in the literal meaning. Even fewer full sentences were given as examples for *surmata*.

a) Functions

Surmata can head kernel clauses according to Finnish dictionaries:

> Surmasi kimppuunsa hyökänneen koiran. [Killed the attacking dog.] (NSS)

This is probably the most basic use of the verb, as the dictionaries which only give few examples have this use in them. *Surmata* can also follow auxiliaries:

> Sitruunahappo voi surmata bakteereja. [Citric acid can kill bacteria.] (NSS)

b) Inflections

Like *tappaa*, *surmata* has numerous inflected forms:

> Vahingonlaukaus *surmasi* miehen. [Accidental shot killed a man.] (SSKS)
>
> Hänet löydettiin raa'asti *surmattuna*. [He was found cruelly killed.] (KTS)
>
> Tottumaton oksien katkoja helposti *surmaa* puun. [An inexperienced person lopping the branches can easily kill the tree.] (NSS)

It should be pointed out that the *surmata* entries are usually much shorter than the *tappaa* entries, and so they also have fewer examples, and fewer inflectional forms are represented.

c)Dependents

Unlike for *tappaa*, there are no clear intransitive uses given in dictionaries for *surmata*. *Suomi-Englanti Suursanakirja* (s.v. *kill* 1) actually only gives *surmata* as an equivalent for transitive *kill,* further reinforcing the idea that it can only be used transitively. When used in transitive constructions, the objects of *surmata* vary, with the restriction that the literal sense usually takes animate objects:

> Surmasi *käärmeen* kepillä. [Killed a snake with a stick.] (KTS)
>
> Tykkituli surmasi kaksi *miestä*. [The cannon fire killed two men.] (NSS)

There is also the example above with a tree as the object of *surmata* in its literal sense:

> Tottumaton oksien katkoja helposti *surmaa* puun. [An inexperienced person lopping the branches can easily kill the tree.] (NSS)

Nevertheless, NSS states that this is a rare use. The objects of *surmata* in a figurative sense seem to vary more, as exemplified by the two sentences already cited in chapter 4.1.3.:

> Surmata joku katseellaan. [Kill someone with your eyes.] (SSKS)
>
> Surmasi haluttomuudellaan vanhempiensa toiveet. [His unwillingness killed his parents' hopes.] (NSS)

Here, '*joku*' refers to an animate object (a human being) while '*vanhempiensa toiveet*' is inanimate and abstract.

5. Contrastive analysis based on the dictionaries

The structure of the following table is based on Lado's model (1957, 67-69) explained in chapter 3.3 where I discussed the methods of this study. Based on the dictionary study, *kill* was found to occur in the structures in the first columns. The second column contains the example sentences in English,

and the third and fourth ones equivalent sentences in Finnish, if one can be given. The sentences themselves have been made up as the sentences in the corpus would have been too long to fit the table and the corpus data will also be analysed only starting in the next chapter.

kill (literal)	*kill*	*tappaa*	*surmata*
Intransitive	*Heroin can kill.*	*Heroiini voi tappaa.*	*Heroiini voi surmata.*
Transitive (+animate object)	*Cain killed Abel.*	*Kain tappoi Abelin*	*Kain surmasi Abelin.*
Transitive (+inanimate object)	*It killed the weeds.*	*Se tappoi rikkaruohot.*	*Se surmasi rikkaruohot.*

kill (figurative)	*kill*	*tappaa*	*surmata*
Intransitive	*If looks could kill...*	*Jos katseet voisivat tappaa...*	*Jos katseet voisivat surmata...*
Transitive (+animate object)	*This hike is killing me.*	*Tämä vaellus tappaa minut.*	*Tämä vaellus surmaa minut.*
Transitive (+inanimate object)	*It killed my chances.*	*Se tappoi mahdollisuuteni.*	*Se surmasi mahdollisuuteni.*

Table 1.

The first division, based on the meanings of *kill*, was into the figurative and literal. This on its own does not help differentiating between *tappaa* and *surmata,* as both can occur with literal and figurative meanings. Further division was based on the types of objects *kill* takes, as was mentioned in chapter 4.2.1 c), objects are a verb-specific type of dependent, and *kill* takes a wide range of objects and also occurs intransitively. This division proved somewhat more interesting, as while *tappaa* was found with all the same kinds of objects in the dictionaries as *kill* was, *surmata* had a narrower range. For example, intransitive *surmata* was missing from the dictionaries all together (see chapter 4.2.3 c)). However, as the *surmata* entries in all dictionaries used were much shorter than the *tappaa* entries, this can simply be a case of only including the more prominent uses. Thus, in table 1., none of the *surmata* sentences can be said to be totally ungrammatical, even though some of them might sound more peculiar than the corresponding *tappaa* sentences. Perhaps the only one where the two are equally acceptable is the second one, literal + animate object. But that is just

my opinion as a native speaker of Finnish, and actual study would require a much larger group of people to evaluate the sentences, which is unfortunately beyond the scope of this study.

After looking at the dictionary entries for *kill, tappaa* and *surmata*, it has become clear that it is not possible to assign a specific Finnish equivalent to a specific meaning of *kill*. Because it is often claimed (see for example Palmer 1976, 60; Cruse 1986, 270; Edmonds and Hirst 2002, 3) that no absolute synonyms exist, as languages strive to be efficient, there must be certain criteria that make a language-user choose between alternatives like *tappaa* and *surmata*. As these criteria could not be found in dictionaries, further research is clearly necessary. For the purposes of the present study this means carrying out a corpus analysis to discover the patterns that occur with the different equivalents of *kill* in actual language usage.

6. Corpus analysis

Chapter 3.2 explained how the corpus data was gathered. After those preliminary steps, the first part of the actual analysis was to categorize the sentence pairs according to the equivalent of *kill* or *tappaa* that was found. All the percentages in the following tables have been rounded to the nearest whole number, denoted by ”~”. If the symbol is missing, that means the figure was not rounded. The results of this categorization were the following:

Equivalent of *kill*	*tappaa*	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Number of sentences with the equivalent	152 ~64%	34 ~14%	18 ~7%	8 ~3%	8 ~3%	21 ~9%	241 100%

Table 2.

These categories do not include only the word in the title, but also other words derived from it or phrases based on it. Thus, the *kuolla* category for example also includes sentences with *kuolema* or *kuollut*, and *henki* refers to phrases such as *päästä hengestään* or *ottaa hengiltä*. These broader categories were used because of number of sentences in each category would have been too small for analysis otherwise. This is the categorization that will be used for most of this study. However, I

have included a list of all the specific words found in Appendix 1.

Based on this, *tappaa* is clearly the most common Finnish equivalent of *kill*. *Kill* could simply be translated as *tappaa*, and that would be correct in more than half of the cases. If further research shows no criteria where the other equivalents are percentually (because in absolute numbers, it is very unlikely they will surpass *tappaa*) more common than *tappaa*, this might be the best solution. However, a translation that does not work ~35% percent of the time is not a very satisfactory, meaning that this further research is necessary.

It is somewhat interesting to see *kuolla* as the third most common equivalent, as the relationship of *kill* and *kuolla* is not as straightforward as *kill* and *tappaa* or *surmata*. The fact that it is so common might be explained by treating *tappaa* as a causative of *kuolla*, a relation that is similar to that between English verbs *kill* and *die*, i.e. *kill* has the meaning 'cause to die'. This will be further examined in chapter 6.2 below.

The category titled ”Other” includes all those words or constructions that did not occur often enough to warrant a closer examination. Some of them did occur more than once, though, as seen on the following table:

murhata/*itsemurha*	*ampua*	omission
3	5	4

Table 3.

Mostly the section includes sentence pairs where the equivalent chosen is specific in the method of killing, such as the *ampua* ones or is a less common synonym of one of the words in the other categories. The following are examples of these types:

> Laulu kertoi miehestä, joka **teurastettiin** rakkautensa tähden kuin lammas ja nyljettiin kuin kala.
> (LAN 14:12)
> The song told of a man who for his love's sake was **killed** like a lamb and skinned like a fish.

Kajaanissa **menehtyi** nuori mies liikenneonnettomuudessa.
(YLE8 5:1:3)
In Kajaani, a young man was **killed** in a traffic accident.

Equivalent of *tappaa*	*kill*	Other	Total
Number of sentences with the equivalent	154 ~89%	20 ~11%	174 100%

Table 4.

On the Finnish side, things are even more straightforward. *Kill* is the equivalent of *tappaa* in 89 percent of the cases and no other equivalent among the others was common enough to examine on its own. *Slay* appeared three times; omission, *slaughter*, *destroy* and *wipe out* twice.[2] Of the rest, most were of the method-specific type mentioned above, such as the following:

Olivat **tappaneet** sata sikaa, jotka näyttivät vielä olevan niin hyvässä lihassa, ettei niitä edes olisi tarvinnut lihottaa.
(MER 3:223)
A hundred pigs had been **shot**, pigs in such a pink of condition that they were ripe for the butcher's knife without the waste of a bucket of feed.

This means that solutions like the one suggested above, about always choosing the most common word, would work even better, and is likely to be to only one possible, as it is difficult to do statistical analysis of the other equivalents based on a couple of examples alone. Based on this, I have decided not to go through the same type of in-depth analysis for *tappaa* as I will use for *kill* in the following subchapters, as it would most likely reveal nothing.

The correlation of *kill=tappaa* is not that straightforward, however. The simple fact that these other equivalents for both were found in the data means that they are part of their respective languages, and also have some aspect in them that makes them different from *kill* and *tappaa*. Reducing everything to just those two most basic words would mean simplifying the languages,

2 I had originally thought to examine the near synonyms of *kill* alongside it, but these, such as *slay* and *slaughter* proved to be so rare in TamBiC (less than ten appearances each) that I was unable to follow this plan.

taking something out of them.

In the following sub-chapters I will examine various criteria which have potential for differentiating between the equivalents. Some were chosen because they clearly fit into Huddleston's framework (e.g. inflections), others because the dictionary analysis mentioned them (e.g. Literal and figurative uses), and the rest because they formed patterns that caught my eye while doing preliminary examination of the data (e.g. passives). The order of the sub-chapters goes from those small in scope and easily recognisable to those where the larger context need to be examined to categorize the data. Incidentally this also means that those chapters where the phenomenon examined fits into Huddleston's framework come first. Each chapter begins with three tables, the first one giving out the distribution of the occurrences in absolute numbers, the second one giving percentages based on the occurrences of the phenomena discussed with a specific equivalent, and the third one giving percentages based on the occurrences of the different equivalents inside a specific phenomenon. The first tables can be read in either direction, but the second tables should be read from top to bottom while the third tables should be read from left to right when it comes to the totals.

6.1. Inflections

Inflections are one of Huddleston's criteria for defining word classes, as was mentioned in chapter 4.2, and make a straightforward starting point for this analysis chapter, as *kill* can occur in all six of the English inflectional verb forms (as shown in chapter 4.2.1 b). Examples of all of these forms were also found in the TamBiC data, though some were much rarer than the others. This chapter is divided so that the analysis focuses on one inflected form at a time, each in their separate section, with a short discussion of the distribution of the form in general at the beginning.

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Base form	56	7	1	3	0	7	74
General present tense	6	2	0	0	0	1	9
3rd person -s	11	7	2	1	0	0	21
Past tense -ed	21	2	2	2	0	3	30
Past participle -ed	39	16	11	2	8	8	84
Present participle -ing	19	0	2	0	0	2	23
Total	152	34	18	8	8	21	241

Table 5.1 – Inflections

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
Base form	~37%	~21%	~6%	~38%	0%	~33%
General present tense	~4%	~6%	0%	0%	0%	~5%
3rd person -s	~7%	~21%	~11%	~13%	0%	0%
Past tense -ed	~14%	~6%	~11%	~25%	0%	~14%
Past participle -ed	~26%	~47%	~61%	~25%	100%	~38%
Present participle -ing	~13%	0%	~11%	0%	0%	~10%
Total	100%	100%	100%	100%	100%	100%

table 5.2 – Percentages according to equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Base form	~76%	~9%	~1%	~4%	0%	~9%	100%
General present tense	~67%	~22%	0%	0%	0%	~11%	100%
3rd person -s	~52%	~33%	~10%	~5%	0%	0%	100%
Past tense -ed	~70%	~7%	~7%	~7%	0%	~10%	100%
Past participle -ed	~46%	~19%	~13%	~2%	~10%	~10%	100%
Present participle -ing	~85%	0%	~8%	0%	0%	~8%	100%

table 5.3 – Percentages according to phenomenon

The two most common forms *kill* was found in in the data were the base form (74 instances) and the past participle (84 instances). The least common by a wide margin was the general present tense, with only 9 instances (table 5.1). 3rd person, past tense and present participle are closer to each other, with 21, 30 and 23 instances respectively. In general, based on this data it seems that *kill* is more commonly used when discussing either potential killing or killing that has already happened in the past than killing that is currently going on. The following give examples of all three cases.

Potential: In a desperate situation, an attacker might **kill** a child or hack off a limb in a bid to "silence" a microchip.
(ST2 9:2:68)
Epätoivoisessa tilanteessa kaappaaja saattaisi **tappaa** lapsen tai hakata irti lapsen raajan eliminoidakseen mikrosirun.

Past: After his uncle was **killed** by Russian troops in June 2001, Barayev took over the leadership of the 300-strong group.
(ST2 10:3:81)
Kun Venäjän joukot olivat **tappaneet** hänen setänsä kesäkuussa 2001, Barajev nousi 300-henkisen joukon johtajaksi.

On-going: But such omnivorous birds **kill** with blows of their beaks.
(ATT 4:110)
Nämä kaikkiruokaiset linnut **tappavat** saaliinsa nokan iskuilla.

(a) Base form

The base form occurred in various structures, the most common being with different types of auxiliary verbs, such as *will*/*would*, *can*/*could* and even *shall*, but there were also *to*-infinitives, negatives and imperatives. The distribution of the various auxiliaries would have been an interesting additional category for this section, but unfortunately the data set was too small for this. The following are examples of *kill* used in its base for in the structures mentioned above.

Auxiliary: According to Jacques Lameloise, another three-star chef, Loiseau told him that if he lost a star, he would **kill** himself.
(ST3 3:3:41)
Kolmen tähden kokki Jacques Lameloise kertoo, että Loiseau oli sanonut hänelle **tappavansa** itsensä jos menettäisi yhden tähden

To-infinitive: Kun kerran laukasee, niin siinä on oltava **tappamisen** meininki.
(LIN 9:5:29)
When you pull the trigger, do it to **kill**.

Negative: "You didn't **kill** Joanna," said Bethany, "only her boyfriend, so what are you going on about?"
(WEL 16:65)
"Et sinä **surmannut** Joannaa", Bethany sanoi, "vain hänen poikaystävänsä, niin että mitä sinä oikein jauhat?"

Imperative: "**Kill** the pig! Cut his throat! **Kill** the pig! Bash him in!"
(GLD 7:146)
"Sika **tappakaa**! Kurkku leikatkaa! Sika **tappakaa**! Kurkku poikki vaan!"

Looking at the different equivalents, the prevalence of *tappaa* as an equivalent is clear from the start. None of the others come even close to it in absolute figures or percentages when *kill* is in the base form, with slightly over three quarters of the sentence pairs found being ones with *tappaa* (table 5.3). In fact this is more than *tappaa*'s proportion of all the examples (~64%, table 2), which would mean that *kill* being in the base form would be a case for even stronger preference for *tappaa* than generally.

(b) General present tense

As mentioned above, the general present tense was the least common inflectional form of *kill* in the data. With only nine examples, it is difficult to say anything concrete about it, as all percentages would change greatly with the addition of even one more example.

Even in this small amount of examples, two distinct categories emerged. One was the news-paper headline, other the ”present habit”-type of use (Leech & Svartvik 2002, 67). The latter of these occurring in the data is somewhat unnerving, but most of the occurrences actually refer to animal behaviour. The following are examples of these two types of use.

Headline: Horrified Visitors See Bears **Kill** Zoo Keeper
(REU 4:5:1)
Kauhistuneet vierailijat näkivät karhun **tappavan** hoitajansa

Habit: But such omnivorous birds **kill** with blows of their beaks.
(ATT 4:110)
Nämä kaikkiruokaiset linnut **tappavat** saaliinsa nokan iskuilla.

(c) 3rd person singular

It is quite interesting that the 3rd person singular present tense was more common than the general

present tense, though it was still the second rarest form. Perhaps killing is not one of those things one talks about doing oneself, and it also usually takes only one person or creature to do it. Chapter 6.4.2 will take a closer look at the types of subjects *kill* actually takes in the corpus data. The situations where *kills* occurred were quite varied and no new clear-cut categories emerged here like they did with the general present tense. The ”present habit” examples were still there, even in cases resembling those of general present tense, only this time with singular subjects. For example, compare the following to the second example sentence in section b) above:

There it **kills** its victim by striking its head sharply against its perch.
(ATT 3:363)
Se **tappaa** uhrinsa iskemällä sen päätä lujasti oksaa vasten.

Using the ~64% of *tappaa* in all the data as the baseline, it was also the first of the inflected forms where *tappaa* was slightly less common with ~52% (table 5.3). Of the other equivalents, *surmata* was the one which saw a certain rise in its frequency, as a third of the 3rd person singular occurrences were with this equivalent (table 5.3) and its baseline is ~14% (table 2). On the other hand, 3rd person singular was completely missing from the ”other” category. It is difficult to think of any particular reason for either of these cases based on the data. Some of the cases where *tappaa* and *surmata* were used were very similar to each other, such as the following.

At the same time they will also publish the genome of the parasite itself – known as plasmodium – which **kills** by infecting and destroying red blood cells.
(ST2 9:6:43)
Samaan aikaan he julkistavat myös itse loisen, plasmodiumin, genomin. Loinen **tappaa** tarttumalla punaisiin verisoluihin ja tuhoamalla niitä.

He wants to use the embryo to carry out research and seek a cure for motor neurone disease, which **kills** 100,000 people a year worldwide.
(ST3 4:5:5)
Wilmutin tarkoituksena on tutkia kloonattujen alkioiden avulla vaikeata lihaksiston liikehermosairautta. Hän hakee myös mahdollista hoitoa tähän tautiin, joka **surmaa** maailmassa 100 000 ihmistä vuosittain.

(d) Past tense

The past tense is in the middle of the range of commonness with its 30 occurrences. *Tappaa* was slightly more common with the past tense than the baseline with ~70% of the occurrences being with *tappaa*. The remaining ~30% was spread quite evenly among the other equivalents, with *surmata*, *kuolla* and *henki* all claiming ~7%, and ”other” 10% (table 5.3). However, this does mean that *surmata* was less common here than its baseline of ~14%, and *henki* more common than its baseline of ~3% (table 2) It should be noted that as the *henki* category consists of only 8 examples, even slight changes will change its percentages drastically, and thus not much concrete can be said based on them. Even the slightly more common *surmata* and *kuolla* had two occurrences each, so no categories or tendencies can be based on them.

The sentences themselves were very straightforward, which might explain the prevalence of *tappaa*, the most basic of equivalents. Perhaps interesting is the fact two out of three occurrences in the ”other” category had *ampua* as the equivalent. However, as table 3 reveals, *ampua* is the most common of the equivalents in the category, occurring five times, this is not an especially strong connection. It cannot even be said that the phrase ”shot and killed”, which is used in both of these cases is always the equivalent of *ampua*, as it also occurs with one of the *surmata* examples:

> Illinois police shot and **killed** a 400-pound Bengal tiger that escaped from its owner at a roadside truck stop near a residential area on Saturday.
> (REU 3:13:3)
> Illinoisin poliisi ampui ja **surmasi** sunnuntaina 180-kiloisen Bengalin tiikerin, joka oli karannut omistajaltaan lähellä asuinaluetta sijanneelta huoltoasemalta.
>
> - Mun mies **ammuttiin** kolme vuotta sitten Ähtärin keskustassa,
> (LKS 3:3:3)
> "They shot and **killed** my husband three years ago, in downtown Ahtari.

(e) Past participle

Past participle was the most common of *kill*'s inflectional forms with 84 occurrences and all of the

different equivalents occurred with it (table 5.1) It was the most common form with *surmata, kuolla, kaatua* and ”other”, and second most common with *tappaa* and *henki* (table 5.2). Interestingly enough, it was also the form where *tappaa* was at its least common among the equivalents, with only ~46% of the occurrences covered by it (table 5.3), which is clearly under the baseline of ~64% This means that the past participle of *kill* is worth close examination, as it can reveal something about the situations when an equivalent other than *tappaa* is used.

Especially noteworthy is the fact that all the instances of *kaatua* occurred with this inflection. Closer look at them reveals that this is because all of them occurred with a passive form of *kill.* Thus the situation will be examined more closely in chapter 6.3, which focuses on the passive. The strong preference *surmata* (~47% of the occurrences were with this form) and *kuolla* (~61%) (table 5.2) also showed for the past participle can be partially explained by their preference for the passive *kill* as well. In their cases it does not cover all of the occurrences, but does a significant amount of them. Out of 16 instances of *surmata* with the past participle, only three of them were with active *kill*. For *kuolla* the figure was 1 out of 11. All in all the figure was 23 out of 84 (figures for the passives are in table 6.1 in chapter 6.2).

Rest of the past participle occurrences were mostly quite straightforward examples of *kill* in an active perfect aspect, mostly in present perfect but there were some cases of past perfect as well. Here are examples of these and the passive as well.

Present perfect: You know, I've never **killed** a man - not like that, not face to face.
(SMI 5:6:124)
Enhän ole ikinä **tappanut** ketään - en sillä tavalla, kasvotusten.

Past perfect: Wallace called the 911 police emergency line from a pay phone and said she had **killed** her young son, King said.
(REU 2:2:15)
Wallace soitti poliisin hätänumeroon kolikkopuhelimesta ja kertoi **tappaneensa** nuorimman poikansa, King kertoi.

Passive: A 27-year-old man was **killed** swimming off North Carolina's Outer Banks barrier islands and his companion was severely injured.
(REU 4:14:10)
27-vuotias mies **sai surmansa** uidessaan Pohjois-Carolinan Outer Banks -rajasaarilla. Hänen kumppaninsa loukkaantui vakavasti.

(f) Present participle

Last of the inflectional forms is the present participle *killing*. This form was in the rarer end of the scale with 23 occurrences (table 5.1). It only occurred with *tappaa*, *kuolla* and ”other”, and even though the other two did exist, *tappaa* claimed 19 out of those 23 occurrences, while *kuolla* and ”other” both had two (table 5.1). In fact, present participle was the inflectional form with highest percentage of *tappaa*, ~85% (table 5.3) Interesting is the total lack of *surmata*, as as the second most common equivalent, one would expect it to appear if the less common *kuolla* does as well.

The present participle is an interesting and somewhat difficult form when examining verbs solely, as it can be used in various ways, some of which are more clearly verbal than others. Recognising the form itself is easy, determining exactly how it has been used more difficult. In this data, the present participle was used in various ways. Perhaps rarest was the most verbal one, progressive ”be + killing”, of which only four examples could be found, three past and one present, all with *tappaa*. The following are examples of this use:

Apina **tappoi** täitä turkistaan.
(K-M 1:62)
"The monkey was **killing** fleas in its fur."

- Kaverit... auttamaan! Minut **tapetaan**, mylvi mies Törmälän isännän raskaan ruhon puristuksessa ja yritti tempoilla vapaaksi.
(PTL 6:3:461)
"Buddies... Help! They're **killing** me!" bellowed the man trying to free himself from under the heavy bulk.

Gerunds, which I have decided to include in this study to maximise the amount of data despite their status as verbal forms being open to discussion, were more common, covering the remaining 19

instances. The fact that gerunds can be perceived as non-verbs can also be seen in the Finnish equivalents that have been used for *killing*. As seen in Appendix 1, the equivalent categories include other words closely related to the one used for the title of the category. The table in the appendix lists ten instances of *tappaminen* and two instances of *tappo*, both of which are nouns. Out of these twelve, half were used with *killing*, which as mentioned, is a relatively rare form. Both the ”other” occurrences of *killing* also had a noun equivalent, *itsemurha* and *hirvenkaato*. On the other hand, both the *kuolla* equivalents were exactly that, using the past form *kuoli*, and several of the *tappaa* instances used the verb in some form as well. This shows that the case is not straightforward, and in my opinion also that gerunds have their place in the data of this study. The following are examples of both types, noun and verb.

Noun
Military intelligence sources said **killing** the rebel leader had become a personal quest for Putin...
(ST2 12:2:13)
Sotilastiedustelulähteet sanoivat, että kapinallisjohtajan **tappamisesta** on tullut Putinille henkilökohtainen tavoite...

"Hopefully, people contemplating suicide will listen to our music and see our posters and get diverted from **killing** themselves at the stations," Banerjee said.
(REU 2:13:10)
"Toivottavasti itsemurhaa suunnittelevat kuulisivat musiikkimme ja näkisivät julisteemme ja muuttaisivat mielensä **itsemurhan** suhteen", Banerjee sanoi.

Verb
seizing the black, two-foot-long reptiles in its talons and eventually, often after a long struggle, **killing** them by tearing them apart with its beak.
(ATT 4:123)
Se sieppaa mustia, yli puolen metrin mittaisia liskoja kynsiinsä **tappaen** ne - usein pitkän ottelun päätteeksi - repimällä ne nokallaan palasiksi.

Kolme ihmistä **kuoli** perjantaina Taipalsaarella Etelä-Karjalassa pienkoneen pudottua Saimaaseen.
(HS8 8:23:2)
A light plane crashed into Lake Saimaa on Friday afternoon, **killing** all three persons on board. The accident happened in Taipalsaari, close to Lappeenranta.

6.2 Passive

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Active	130	21	8	6	0	15	180
Passive (all)	22	13	10	2	8	6	61
Be-passive	16	10	7	1	6	4	44
Get-passive	4	2	1	1	1	2	11
Unexpressed verb-passive	2	1	2	0	1	0	6
Total	152	34	18	8	8	21	241

table 6.1 - Actives and passives

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
Active	~86%	~62%	~44%	75%	0%	~71%
Passive (all)	~14%	~38%	~56%	25%	100%	~29%
Be-passive	~11%	~29%	~39%	~13%	~75%	~19%
Get-passive	~3%	~6%	~6%	~13%	~13%	~10%
Unexpressed verb-passive	~1%	~3%	~11%	0%	~13%	0%
Total	100%	100%	100%	100%	100%	100%

Table 6.2 - Percentages according to the equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Active	~72%	~12%	~4%	~3%	0%	~8%	100%
Passive (all)	~36%	~21%	~16%	~3%	~13%	~10%	100%
Be-passive	~36%	~23%	~16%	~2%	~14%	~9%	100%
Get-passive	~36%	~18%	~9%	~9%	~9%	~18%	100%
Unexpressed verb-passive	~33%	~17%	~33%	0%	~17%	0%	100%

Table 6.3 – Percentages according to phenomenon

While *tappaa* is still the most common equivalent in absolute numbers, the others are not that far behind when it comes to passives. Going equivalent by equivalent, *tappaa* is in fact the one with lowest percentage of passives (~14%), the next one (*henki* at 25%) (table 6.2) having almost twice that amount. *Kuolla* was used with the passive in more than half of its occurrences (table 6.2). This

points to passivity being one of those cases where an equivalents other than *tappaa* should be given more weight when choosing an equivalent. The equivalents besides *tappaa* (except those in the ”other” category) claimed slightly smaller portions of the passives with *get*-passive than the *be*-passive (table 6.3). This could be due to the fact that *get*-passive is the more informal of the two (Leech&Svartvik 2002, 346). As was mentioned in chapter 4.1.3 above, *tappaa* is considered more coarse than *surmata*, meaning that it is more likely to be used in informal language.

The sentence pairs in the *surmata* category with passive *kill* are somewhat bifurcated. If the English sentence has an agent, such as in the following example, the verb *surmata* is used in the active, with the agent as the subject.

> mutta pohjoisen ilmastossa käärmeitä **surmaa** eniten pakkanen.
> (TRA I10:9)
> but in northerly climates snakes are most often **killed** by the cold weather.

This particular sentence could not even be translated into Finnish with a passive, as as Finnish passive sentences do not allow inanimate agents *(Iso Suomen Kielioppi,* internet-version=VISK §995). Thus, **käärmeitä surmataan eniten pakkasen toimesta* would be a completely ungrammatical. Sentences using this construction were a minority in the data, however. More common were the cases where the equivalent of passive *kill* was the construction *saada surmansa*. Here is an example of this sentence type:

> As the war ended, Hassan spread a rumour that he had been **killed** in the attack on Baghdad airport.
> (ST3 6:11:13)
> Sodan loputtua Hassan levitti huhua, jonka mukaan hän olisi **saanut surmansa** pääkaupungin lentokentälle tehdyssä hyökkäyksessä.

These sentences are not actually examples of the verb *surmata* in use, as *surma* is a noun rather than a verb. Of the *surmata* category, 8 of the occurrences actually used *surma* instead of *surmata* (see

Appendix 2). But, as referred to earlier, Finnish passive constructions are quite limited in their use. Because of this, *saada surmansa* can be seen as an equivalent of the passive form of *surmata*. It is certainly used in this way based on my data. This close connection is one of the reasons for grouping *surmata* and *saada surmansa* into one category in this study. Krzeszowski (1990, 165) brings up the thought-provoking point that on syntactic or sentence level, English passives do not have Finnish equivalents. This, however does not mean that there does not exist a word level or pragmatic equivalence, and I think the relationship between *kill* and *saada surmansa* is good example of this.

Kuolla had the second highest percentage of passives among the equivalents. As was stated at the beginning of chapter 6, there exists a causative relationship between *kuolla* and *tappaa*, and the English *kill* and *die*. While dying does not imply killing, killing does imply dying, making *die* a hyperonym of *be killed*. This can explain the amount of sentence pairs with *kuolla* (and its derivations) as the equivalent in the data. It is same kind of word choice as those more specific equivalents in the ”other” category (also mentioned at the beginning of chapter 6), except here the word chosen is a more generalized one. *Kuolla* is used very similarly to *saada surmansa* as a way to avoid using *tappaa* or *surmata* in the passive, due to restrictions of the Finnish passive:

> mutta sitten hän oli äkkiä entistä onnellisempi siitä ettei faija ollutkaan **kuollut** metron alle. (JNS 54:30)
> Gradually he began to feel happier than ever that his father hadn't been **killed** in the underground accident.

Deciding between *kuolla* and *saada surmansa* seems to depend on other factors, however, such as text types, which will be discussed later in chapter 6.5.

Kaatua is an especially interesting case, as it does not appear with active *kill* at all. Based on this it seems that *kaatua* can only be an equivalent of passive *kill*. The eight *kaatua* sentence pairs actually form a very uniform category, as this passivity is not the only aspect they all have in

common. Them all having the same verb form of *kill* (past participle) is of course because of the use of passive, but further chapters will show that they do not differ in other aspects that much as well. Being an equivalent of *kill* is not a very typical use of *kaatua*, but significantly, all eight have a war context, such as the following:

> Heistä yksi **kaatui** 1964 ja vuoden 1974 sodassa 17 suomalaista haavoittui.
> (TAR 1964:6)
> One of them was **killed** in 1964, and 17 were wounded in the 1974 war.

In fact, *Englanti-Suomi Suursanakirja* supports the idea that *kaatua* is the equivalent of choice when *kill* is in passive and the context is war (or military in general) by stating (s.v. 1 *kill*) ”be [kill]ed … (sot) kaatua”. There is one sentence pair in the ”other” category which is similar to the *kaatua* ones but uses a different equivalent, however.

> Pari poikaa on **menny** ja kolme vietihin heikos kunnos kenttäsairaalahan.
> (LIN 10:4:126)
> Two lads were **killed** and three taken to the hospital in bad shape.

This is clearly a dialectal or colloquial use, and because of this something that is quite limited in its use. *Kaatunut* would work on the example sentence quite well in place of *menny*, which means that this sentence pair does not negate the earlier claim of *kill* in the passive plus war equals *kaatua.*

6.3 Semantic objects and subjects

It should be noted that here the terms ”object” and ”subject” refer to the semantic roles, not the grammatical ones. In the statistics below, ”subject” refers to the person/animal/thing that does the killing (semantically known as the 'agent'), and ”object” to the person/animal/thing that is killed (semantically, the 'patient'). Thus, for example agents of passive sentences were categorized as ”subjects” and the grammatical subjects of those sentences as objects. This was done because the

thing I wanted to examine was the relation of the type of killer and killed to the equivalents, most importantly, whether there is any difference between human and non-human objects and subjects. This is mostly based on the mention of *surmata* being ”less coarse” than *tappaa* in NSS (s.v. *surmata*), the assumption being that more coarse words are used with non-humans, whether objects or subjects, which would mean that *surmata* is the most likely when both the killer and the killed are human. While it might be interesting to cross-reference between the objects and subjects, that is beyond the scope of this study and because of this the two will be handled in separate subchapters.

6.3.1 Objects

As noted in subchapter 4.2.1 b), *kill* can be used both transitively and intransitively. In the following tables, the intransitive cases are handled as if they were one category of objects, with the object being ”none” or ”zero”. This was done to make the figures comparable with those of other tables, as taking the intransitives out would have meant different numbers for the totals.

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Intransitive	10	1	0	0	0	0	11
Human	88	18	16	6	8	17	153
Animal	42	9	2	2	0	3	58
Inanimate	7	6	0	0	0	1	14
Unclear	5	0	0	0	0	0	5
Total	152	34	18	8	8	21	241

Table 7.1 - Objects

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
Intransitive	~8%	~ 3%	0%	0%	0%	0%
Human	~58%	~53%	~89%	75%	100%	~81%
Animal	~28%	~26%	~11%	25%	0%	~14%
Inanimate	~5%	~18%	0%	0%	0%	~5%
Unclear	~3%	0%	0%	0%	0%	0%
Total	100%	100%	100%	100%	100%	100%

Table 7.2 - Percentages according to the equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Intransitive	~91%	~9%	0%	0%	0%	0%	100%
Human	~58%	~12%	~10%	~4%	~5%	~11%	100%
Animal	~73%	~15%	~ 3%	~3%	0%	~5%	100%
Inanimate	50%	~43%	0%	0%	0%	~7%	100%
Unclear	100%	0%	0%	0%	0%	0%	100%

Table 7.3 - Percentages according to the type of object

There are no object categories where the absolute numbers of any other equivalent than *tappaa* would exceed its. Table 7.3 does however reveal that the percentage of *tappaa* does vary. It is at its most common in those cases where the object is unclear, as all of those cases in the data occurred with *tappaa*. There were only five of them (table 7.1), four of which were truly without any way of determining the type of object and one where there was no clear way to choose between two of the categories. Cases such as the ones below, where there is unclarity because of the use of personal pronouns or a descriptive phrase, would probably be quite easy to determine if the whole text or even a paragraph was given, and thus would not be problems in most cases, but are in some corpus-based studies where the larger context is not provided.

"**Kill** him! **Kill** him!"
(GLD 7:141)
"**Tappakaa**! **Tappakaa**!"

Ultimately, however, it did die. I forget what **killed** it.
(WLD 8:218)
Mutta vihdoin se kuitenkin kuoli. En muista enää, mikä sen **tappoi**.

Sieltä sen koukkunokan **tappaisi**, jos olisi vähänkään reikärautaa mukana, hän pahoitteli.
(PTL 7:2:41)
"I could **kill** that hook-beak if I had any kind of weapon," he lamented.

Ei sunkaan tässä kukaan rupee sormi suussa **tappamistansa** vartoomaan.
(LIN 9:3:71)
Who do you think's going to stand with his finger in his mouth waiting to get **killed**?

Cases such as the following would be more difficult, however:

Vihan ja pettymyksen tunne vain yltyi. Hän hyppäsi luotikasan päälle ja polki niitä kuin olisi **tappanut** käärmettä.
(PTL 7:1:127)
In mounting rage and disappointment, he jumped on the pile of arrows and trampled them as if he were **killing** a snake.

Here the actual object is a thing, but because of a metaphor, it could also be categorized as an animal. Luckily, cases such as this one are quite rare. It is however a good example of the problems semantics bring into a mostly quantitative study such as this one. They do not lend themselves as easily to categorization as more syntactic things such as inflections do.

The second most common category for *tappaa* was the intransitive. Again, the number of occurrences was relatively small, with only 11 cases in total, 10 of which were of *tappaa*, one of *surmata* (table 7.1). When dealing with such small numbers, it is difficult to say whether this can be taken as conclusive proof that the other equivalents could not be used with intransitive *kill* at all, or whether it is just a case of rare phenomenon not appearing in this particular data with them. Chapter 4.2.3 c) does support the former for *surmata*, as the Finnish dictionaries used did not provide any intransitive examples of it, but the fact that it was the only non-*tappaa* equivalent used intransitively counters this. Here are examples of both *tappaa* and *surmata* used intransitively.

The specialist hunters - owls and eagles, hawks and falcons - **kill** with their talons.
(ATT 4:111)
Mutta saalistukseen erikoistuneet linnut - pöllöt, haukat ja kotkat - **tappavat** yleensä kynsillään.

Kohtalonsa täyttäjinä nämä miehet **surmaavat**, hylkäävät, jopa kuolevat, kunhan heidän määrittelemänsä oikeus on tapahtunut.
(G-K 34:11)
Fulfilling their fates these men **kill**, reject, even die, just to satisfy their sense of justice. | General present

The fact that passive sentences are more common with the equivalents other than *tappaa* (chapter 6.2) should probably also be taken into account. Passive sentences have a semantic object by necessity, and thus cannot be intransitive. This would mean that, for example, *kaatua* could never be the equivalent of intransitive *kill*, as analysis in chapter 6.2 revealed that it only occurs with passive *kill*. All of this means that intransitivity is one of those cases where even more weight should be given to choosing *tappaa* over other equivalents than usually.

Going in this descending order, next category is sentences with animal objects, with ~73% of them occurring with *tappaa* (table 7.3). This is in line with the supposition given at the beginning of this chapter about *tappaa* being more common with non-human objects. However, if going by equivalent instead of object type, *tappaa* and *surmata* have percentages quite close to each other in this category, with 28% of *tappaa* and 26% of *surmata* occurring with animal objects (table 7.1), which is evidence for the opposite. The types of animals varied quite a lot, from ” black, two-foot-long reptiles” to ” a Saimaa ringed seal pup”, but mostly comprised of mammals, birds and a surprising amount of fish. This latter is explained by most of the sentences coming from the same text, Ernest Hemingway's *Old Man and the Sea*, which discusses the killing of a fish at length. This text is an interesting case as it is responsible for two of the nine *surmata* occurrences and the single *henki* occurrence in this category, as well as nine of the 43 *tappaa* occurrences, all of which refer to a single entity, a fish. As the text is a novel, some variation is to be expected so that the text does not feel stale, which is an interesting factor to consider for the choice of equivalent, and raises

questions. How close does the previous occurrence of the word have to be for it to affect the choice of equivalent? How big of a role does the text-type play in this choice? Or the amount of times the word is repeated? Unfortunately this kind of frequency study is not the focus of this study, and thus will be left for others.

Some of the instances in the animal category could have been put into the ”unclear” category based solely on the evidence in the sentences themselves, but sometimes it was quite easy to deduct for example what a personal pronoun referred to. This is of course helped by there being different third person pronouns for humans and non-humans. But there are cases where it is not that simple, such as the following:

> If he will jump I can **kill** him. But he stays down forever.
> (HEM 6:93)
> Jos se hyppää, voin **tappaa** sen. Mutta se pysyy ikuisesti syvällä.

Here the pronoun ”he” is used, but the sentence was categorized as ”animal subject” despite of this. This is because I felt that if I knew for any reason what the category should be, I would not categorize anything as ”unclear”. Here the reason were the other sentences that came from the same text and were close to this one. The text is *Old Man and the Sea*, a fishing narrative, and because of those surrounding sentences, it was quite clear that ”him” in the example sentence referred to a fish. I brought this up mostly to show that the used pronoun cannot be taken as an absolute proof of the object's humanity or non-humanity, at least in this type of non-fiction, though it raises the question of how much does the researcher’s previous knowledge affect corpus studies. In this case, the sentence could be categorized based partially on the fact that I knew what *Old Man and the Sea* has as its subject material, where as those cases that which I decided were unclear were it mostly because I was not familiar with the texts.

At ~58% *tappaa* (table 7.3) we come to the most common type of object of *kill*, humans. There is no question about them being the most common objects of *kill*, as there were 154 sentences

in the data with human objects, with animals being the next most common with 60 sentences (table 7.1). There were some problems of the same type as with the ”animal” category in that sometimes pronouns made it difficult to determine the object clearly. For example, there were several cases where the object was ”me”, such as the following:

"So. Are you going to **kill** me, my friend?"
(SMI 5:6:27)
"Selvä. **Tapetaanko** minut, ystäväni?"

These are most likely going to be humans, as nothing else would realistically be capable of thought, but in fiction, there is a chance that the ”me” in the sentence is actually something else, such as a talking animal or some other sentient species. However, in my categorization those would still have fallen into the ”human” category, which I would have then named ”sentient being” or ”humanoid”. This is because these kind of beings would probably fit into this category better than in anything else. Fortunately, none of the source texts in TamBiC have science fiction or fantasy as the text type, and as such are unlikely to include cases like these. This is merely an interesting aspect of categorizing to ponder, though it has no direct influence on the study at hand.

In the ”human” category, *tappaa* and *surmata* are almost as close to each other as in the ”animal” category, with ~58% and ~53% percent of their occurrences respectively (table 7.1). The smaller percentage of *tappaa* in this category is explained by all the other, less common equivalents being much more common with human objects than in any other category. ~89% of *kuolla*, 75% of *henki*, 100% of *kaatua* and ~81% of the ”other” category occurred with human objects (table 7.1). *Kaatua* is a particularly noteworthy case, as it is one of the few times in this study where the percentages show a clear 100% preference for a certain phenomenon, in this case humanoid objects. For example:

Thousands were **killed**, including Evangelia's younger brother, Filon.

(CAL 3:214)
Tuhansia **kaatui**, muiden muassa Evangelian nuorempi veli Filon.

Kaatua has proven to be quite a specific case, as it had a 100% preference for occurring with *kill* in the passive as well (see chapter 6.2). It might be that the conditions for using *kaatua* can actually be given with almost absolute certainty, unlike what seems to be the case with the other equivalents based on the data examined thus far. Unfortunately, *kaatua* remains a marginal case as an equivalent of *tappaa* with only 8 occurrences in the data.

The next most common equivalent with human objects was *kuolla* with ~89%. This is also an intriguing case, as there is no obvious reason why it would be that much more common with human objects than with animal objects, which accounted for the remaining ~11% (table 7.1). The following are examples of both of these cases.

Human object: Malaria **kills** more than 1m people worldwide each year...
(ST2 9:6:44)
Malariaan **kuolee** maailmassa vuosittain yli miljoona ihmistä,

Animal object: As it dives on them, it swings its legs and pelvis forward, and strikes its victims with such force that the blow may **kill** them outright.
(ATT 4:118)
Kun se syöksyy niitä kohti, se kääntää jalkansa ja alavartalonsa eteenpäin ja iskeytyy saaliiseensa sellaisella voimalla, että saalis **kuolee** heti.

In both cases, *tappaa* could easily have been used in place of *kuolla,* but cases resembling the first example are much more common in the data. There might be some psychological explanation here, if for example it makes people more uncomfortable talking about people being killed instead of dying, while it does not matter for animals, or it can simply happen because of some other phenomenon which occurs much more commonly with human objects than the other types. Unfortunately, examining either of these possibilities is beyond the scope of this study.

The final category are inanimates. This category includes objects, concepts and other non-living things, such as the following:

> Nykyään taiteessa on konstruktio, ja luulen, että juuri se käsite on **tappanut** taiteen".
> (SCH 2:47:6)
> "Nowadays there's all this talk about construction, and I believe concepts have **killed** art."

In this category, *tappaa* and *surmata* were close to each other in absolute numbers, with 7 and 6 occurrences respectively (table 6.1), which means that larger percentage of *surmata* (~18%) than *tappaa* (~5%) (table 7.2) occurred with these types of objects. This was also the category where *tappaa* was at its rarest, as only ~50% of this category was of *tappaa* (table 7.3). As besides sentences with *tappaa* and *surmata*, there was only one other sentence in this category (see below), this means that inanimate objects can be taken as a phenomenon where *surmata* can be given stronger preference than usual when choosing an equivalent.

> Pickled herring weds baguettes, and still more sparkling wine is served to **kill** any hint of an eve-related hangover.
> (HKI 2:4:10)
> Suolasilli kruunaa patonginviipaleet, mutta tarjolla on myös lisää kuohuviiniä **siltä varalta, että** joku sattuisi tuntemaan edes vähäisiä krapulaoireita edellisillan juhlimisen jälkeen.

6.3.2. Subjects

For clarity, I will handle the subjects in the same order as the objects in the previous sub-chapter, going from the category where the percentage of *tappaa* is the highest to the one where it is lowest.

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
No subject	21	11	9	3	7	6	57
Human	96	17	1	2	0	12	128
Animal	14	1	1	1	0	2	19
Inanimate	16	4	7	2	1	1	31
Unclear	5	1	0	0	0	0	6
Total	152	34	18	8	8	21	241

Table 8.1 - Subjects

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
No subject	~14%	~32%	50%	~38%	~88%	~29%
Human	~63%	~50%	~6%	25%	0%	~57%
Animal	~9%	~3%	~6%	~13%	0%	~10%
Inanimate	~11%	~12%	~39%	25%	~13%	~5%
Unclear	~3%	~3%	0%	0%	0%	0%
Total	100%	100%	100%	100%	100%	100%

Table 8.2 - Percentages according to the equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
No subject	~37%	~19%	~16%	~ 5%	~12%	~11%	100%
Human	~75%	~13%	~1%	~2%	0%	~9%	100%
Animal	~74%	~5%	~5%	~5%	0%	~11%	100%
Inanimate	~52%	~13%	~23%	~6%	~3%	~3%	100%
Unclear	~83%	~17%	0%	0%	0%	0%	100%

Table 8.3 - Percentages according to the type of subject

The subjects are divided similarly to the objects in that *tappaa* is the most common equivalent in absolute numbers in every category (table 8.1) but there is great variation in the percentages (table 8.3). *Tappaa* is at its most common in the ”unclear” category, where it accounts for five of the six cases, while the remaining one occurred with *surmata* (table 8.1). As there are few enough of these cases, all of them will be given as examples.

Then he was sorry for the great fish that had nothing to eat and his determination to **kill** him never relaxed in his sorrow for him.
(HEM 7:121)
Sitten hänen tuli sääli isoa kalaa, kun se ei ollut saanut mitään syödäkseen, mutta hänen järkähtämätön päätöksensä **tappaa** kala ei tämän säälin vuoksi hetkeksikään horjunut.

"And you thought that if I had a quarter of a chance I'd denounce you as a thought-criminal and get you **killed** off?"
(ORW 2:2:123)
"Ja sitten kuvittelit, että jos tulisi pienikin mahdollisuus, niin ilmiantaisin sinut ajatusrikollisena ja toimittaisin **tapetuksi**, niinkö?"

Siitä saapi käräjillä melkein yhtä kovan tuomion kun miehen **taposta** ja se on oikein.
(PTL 7:2:74)
For that, the court punishes you almost as harshly as for **killing** a man. And it's right to do so.

Ultimately, however, it did die. I forget what **killed** it.
(WLD 8:218)
Mutta vihdoin se kuitenkin kuoli. En muista enää, mikä sen **tappoi**.

in which a woman character was **killed** by having her head shoved into a deep fat fryer.
(GW3 7:4:8)
jaksossa naishahmo **tapettiin** työntämällä hänen päänsä rasvakeittimeen.

It is enough to live on the sea and **kill** our true brothers.
(HEM 7:127)
Riittää, kun joutuu elämään merellä ja **surmaamaan** oikeita veljiään.

In most of these sentences, the problem of determining the subject type arises from the semantic point of view. A grammatical subject is easier to determine, but for example in the last of these sentences, it is difficult to say into which category should ”by having her head shoved into a deep fat fryer”, a gerund, fall. The ”inanimate” category would probably be the closest match, but it is not a perfect fit. In an ideal case, the ”unclear” category would not even exist, but it does show that it is not always possible to categorize every phenomenon fully and without problems.

Next in this descending order are the ”human” and ”animal” categories, with almost the same percentage of *tappaa* sentences (~75% and ~74% respectively, table 8.3). The ”human”

category is by far the most common one with its 128 sentences (table 8.1). Determining the reason for this would be an interesting task that is once again beyond the scope of this study, as the data does not reveal any particular patterns with human subjects. The types of killing humans commit in the data vary quite a lot as exemplified in the following.

The killing can be a crime:

> Wallace called the 911 police emergency line from a pay phone and said she had **killed** her young son, King said.
> (REU 2:2:15)
> Wallace soitti poliisin hätänumeroon kolikkopuhelimesta ja kertoi **tappaneensa** nuorimman poikansa, King kertoi.

Or happen during a war:

> Hyvä jos neljäsosa retkeen osallistuneista onnistui **tappamaan** edes yhden vihollisen.
> (MER 4:324)
> No more than a quarter of the men in a war succeed in **killing** a single enemy.

Or be more of a metaphysical thing:

> "Yes," he cried, "you have **killed** my love.
> (WLD 7:162)
> "Niin", huudahti hän, "sinä olet **tappanut** minun rakkauteni.

Or simply be hunting for food:

> eikä niitä sitä paitsi kannattanut sellaisella helteellä **tappaa** useampia kuin ehti syödä.
> (MER 3:204)
> Besides, the weather was so hot that it wasn't worth **killing** more than one could immediately eat.

However, considering the fact that human objects were also the most common category with 153 sentences (table 8.1), *kill* seems to be most commonly used as a verb in those cases where humans cause the deaths of other humans. Interestingly enough, animal subjects were the second smallest category after the unclear cases with only 14 cases (table 8.1), though one would assume that animals are just as capable of killing as humans. It might be that the source texts in the corpus

simply did not include that many texts where animals are the main focus. Many of the ”animals as subjects” -sentences came from David Attenborough's *The Life of Birds*, a nature documentary, which is one of those cases where animals actually are in the focus.

Thought-provokingly, *tappaa* was the equivalent that had the largest percentage of human subjects, ~63% (table 8.2), which is in contrast with the objects, where it was in the lower end of the scale in this respect. With subjects, *kuolla* and *kaatua* take the last two spots with ~6% and 0% respectively. This is largely explained by them being used in the passive so often, as was discovered in chapter 6.2. Passive sentences only have semantic subjects if they include an agent, and that does not happen in nearly all of the cases. Sentences such as the following are much more common:

> Lashkar-i-Toiba has been accused of mounting a suicide bomb attack on the Indian parliament in December 2001 in which 14 people were **killed**.
> (ST3 3:11:35)
> Lashkar-i-Toibaa on syytetty joulukuussa 2001 Intian parlamenttiin tehdystä itsemurhaiskusta, jossa **kuoli** 14 ihmistä.

On the other hand, deviating from the order of handling the types of subjects, *tappaa* has the smallest percentage in the ”no subject” category with ~14% of sentences with *tappaa* having this type of subject, while *kaatua* leads here with ~88% (table 8.2), or seven out of eight instances (table 8.1), the eight being a passive sentence with an agent:

> ja vain yksi **kaatui** tulessa, vaikka vääpeli ja hänen miehensä olivat ainakin yhtä pahoissa paikoissa kuin muut.
> (MER 4:270)
> only one was **killed** by actual fire, though his platoon was in as exposed positions as any other.

Kuolla also places high in this category, with 50% of its occurrences having no subjects (table 8.2).

The final category is inanimate subjects, where *tappaa* accounts for ~52% of the cases (table 8.3). The subjects of this category include things such as the following:

Pelkkä alkoholin aiheuttama maksasairaus **tappoi** 1 145 ihmistä 2007, lähes kaksinkertaisen joukon kuin 2001.
(HS9 5:3:13)
Alcohol-related liver disease alone **killed** 1,145 people in 2007, nearly double the number for 2001.

mutta pohjoisen ilmastossa käärmeitä **surmaa** eniten pakkanen.
(TRA I10:9)
but in northerly climates snakes are most often **killed** by the cold weather

In general, diseases seemed quite common in this category. Those will be discussed more closely in chapter 6.5, where I examine collocations. *Kuolla* was surprisingly common in this category, with seven of its eighteen occurrences (table 8.1) being in this category. It could be that inanimate things are seen as incapable of actual killing, and thus *kill* becomes *kuolla* or in some cases *aiheuttaa kuolema*, such as in the following sentence pair:

One leak from a refrigerator at a hospital in Cleveland, Ohio, in 1929 **killed** more than a hundred people.
(SCI 10:2:3)
Vuonna 1929 yksi ainoa vuotava kylmälaite **aiheutti** yli sadan ihmisen **kuoleman**

This is of course not always the case, as *tappaa* was still the most common equivalent in this category, and sentence pairs such as this still exist:

"Thus, on present smoking patterns about one third of all the young men in China will eventually be **killed** by tobacco," they said.
(REU 2:10:12)
"Niinpä tupakka **tappaa** jossain vaiheessa kolmanneksen kaikista Kiinan nuorista miehistä, mikäli tämänhetkiset tupakointitottumukset säilyvät", ryhmä sanoi.

In summary, the semantic subjects of *kill* vary greatly, and are not very helpful when choosing an equivalent for it. In cases where equivalents other than *tappaa* emerge as more common than usual, they are often connected to some other phenomenon, such as passive use.

Thus, the subjects are more of a supporting phenomenon than something to base the choice of equivalent on.

6.4 Semantic uses

This chapter is mostly done out of interest in seeing whether the literal and figurative uses are significant in determining the Finnish equivalent of *kill* to be used. Based on the analysis done in chapter 5, this was not the determining factor in deciding between *tappaa* and *surmata*, but while both can be used with both literal and figurative *kill*, it is interesting to see whether they have actually been used and also how the other equivalents are used. It will also hopefully help in answering my research question number 3 about whether the phenomenona affecting the choice of an equivalent are syntactical, semantic or otherwise.

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Literal	138	28	17	7	8	18	216
Figurative	9	4	1	1	0	2	17
Unclear	5	2	0	0	0	1	8
Total	152	34	18	8	8	21	241

Table 9.1- Literal and figurative uses

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
Literal	~91%	~82%	~94%	~88%	100%	~86%
Figurative	~6%	~12%	~6%	~13%	0%	~10%
Unclear	~3%	~6%	0%	0%	0%	~5%
Total	100%	100%	100%	100%	100%	100%

Table 9.2 – Percentages according to the equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Literal	~64%	~13%	~8%	~3%	~4%	~8%	100%
Figurative	~53%	~24%	~6%	~6%	0%	~12%	100%
Unclear	~63%	25%	0%	0%	0%	~13%	100%

Table 9.3 – Percentages according to the phenomenon

The first noticeable thing here is that the literal use of *kill* is much more common than the figurative use, with 216 versus 17 (table 9.1) instances respectively. Even if all the cases where it is unclear whether the meaning is literal or figurative were actually figurative, the situation would not really change, as there are even less of those (8 instances) than there are of the clear-cut figurative use. This means that however the equivalents are spread, their distribution in the figurative category will not be that useful, for in most cases the use will be literal. Here are examples of the literal and figurative uses.

> Literal: Wallace called the 911 police emergency line from a pay phone and said she had **killed** her young son, King said.
> (REU 2:2:15)
> Wallace soitti poliisin hätänumeroon kolikkopuhelimesta ja kertoi **tappaneensa** nuorimman poikansa, King kertoi.
>
> Figurative: "I am glad we do not have to try to **kill** the stars."
> (HEM 7:117)
> "Olen iloinen, ettei meidän tarvitse yrittää **tappaa** tähtiä."

The unclear cases are a bit more difficult. In the case of this study some of them were the result of me not having the whole text to examine, only one sentence or a couple more if specifically looking at the instance more closely in the online version of the TamBiC corpus. Thus, for example the object of *kill* might be completely missing, and objects were one of the main clues into figuring out the type of semantic use. Of course in most cases the whole text is available for examination, but it is difficult to say from how far away in the text one could find the object or some other clarification, thus potentially increasing the amount work required significantly compared to more clear-cut cases. The other reason for unclarity were those cases where hyperbolation was used when threatening someone. It is sometimes difficult to say whether the threat to kill someone is meant to be taken literally or not. Here are example sentences of both of these cases of unclarity.

Missing words: Ultimately, however, it did die. I forget what **killed** it.
(WLD 8:218)
Mutta vihdoin se kuitenkin kuoli. En muista enää, mikä sen **tappoi**.

Hyperbole: And what is more, and have you forgotten; accept this evidence, accept it or I'll **kill** you, that you do not love unloveable me! Unforgivable!
(WEL 11:76)
Ja sitäpaitsi, ja etkö muka muista; hyväksy tämä tosiasia, hyväksy se tai minä **tapan** sinut, että sinä et rakasta minua jota on mahdoton rakastaa! Anteeksiantamatonta!

Furthermore, *tappaa* proves once again to be the most common equivalent, regardless of the category, when looking at the absolute numbers. The percentages are more interesting, as they reveal its relative amount to be around 11% smaller with the figurative use (~53% *tappaa*) than with literal use (~64% *tappaa*) (table 9.3). This means that the in those rare cases where *kill* is used figuratively, equivalents other than *tappaa* could be given more consideration than with the literal use. Out of these other equivalents, *surmata* and *henki* are the ones with highest percentage of figurative instances, ~12% for *surmata*, ~13% for *henki* (table 9.2). It is true that this only means one instance for *henki* and four for *surmata*, but the fact that they could be found at all with this rare use is what is important here. With their severely smaller number of instances than *tappaa*, one might have expected not to find them used with the figurative. On the other hand, *kaatua* does not occur with the figurative use at all, meaning that it could be totally excluded from consideration when *kill* is used figuratively. *Kuolla* is in between these two numbers with it's ~6% (table 9.2), or one example. The ”other” category included two examples of the figurative use (table 9.1), one of them having *murder* as the equivalent, the other going for the complicated construction given below.

Pickled herring weds baguettes, and still more sparkling wine is served to **kill** any hint of an eve-related hangover.
(HKI 2:4:10)
Suolasilli kruunaa patonginviipaleet, mutta tarjolla on myös lisää kuohuviiniä **siltä varalta, että** joku sattuisi tuntemaan edes vähäisiä krapulaoireita edellisillan juhlimisen jälkeen.

This is also a clear example of the fact that the object of *kill* is related to determining its semantic use. A hangover cannot actually be killed, thus the use must be figurative. Even more interestingly, while the subject is not that relevant, the instrument is. Here, it is ”sparkling wine”, which in most cases is not likely to actually kill anything, but that possibility cannot be completely ruled out. They could choke on it, for instance. All of this leads to the conclusion that the dependents and semantics of *kill* are closely connected.

6.5 Text types

Text types are quite different from syntax in the sense that they are determined by the whole text, not just some small part of it, such as one sentence, and they go even beyond that, as they would not even exist if there was not any intertextuality. Baker (2011, 123) calls this ”patterns familiar to the reader. For Stubbs (1996, 3) the meaning of a text is a question of the relationship between the language used, the producer of the text and the person reading it.

The categorization into fiction and non-fiction was done following the TamBic source text list, to which I have also added a more fine division of the non-fiction text types (see Appendix 2).

6.5.1 Fiction & non-fiction

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Fiction	98	13	6	6	6	10	139
Non-fiction	54	21	12	2	2	11	102
Total	152	34	18	8	8	21	241

Table 10.1. – Text types

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
Fiction	~64%	~38%	~33%	75%	75%	48%
Non-fiction	~36%	~62%	~67%	25%	25%	52%
Total	100%	100%	100%	100%	100%	100%

Table 10.2 – Percentages according to the equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
Fiction	~71%	~9%	~4%	~4%	~4%	~7%	100%
Non-fiction	~53%	~21%	~12%	~2%	~2%	~11%	100%

Table 10.3 – Percentages according to the text type

Out of the two, *kill* appeared in fiction more often than in non-fiction. Which category was more common differed from equivalent to equivalent, however. Fiction was more common with *tappaa, henki* and *kaatua*, while non-fiction was more common with *surmata*, *kuolla* and ”other”. Out of these all, clearest was the prevalence of *henki* and *kaatua* in fiction, with both at 75% of their occurrences happening in fiction (table 10.2). Especially in the case of *henki* this might be explained by the fact that the category included various phrases less straightforward than the basic *tappaa*, and fiction tries to avoid repetition more than non-fiction does, resulting in equivalences such as the following:

> This will **kill** him, the old man thought. He can't do this forever.
> (HEM 3:226)
> Tämä kyllä **ottaa** sen **hengiltä**, vanhus tuumi. Se ei voi jatkaa tätä ikuisesti.

Besides these two, *tappaa* showed a strong preference for fiction as well, with ~64% of occurrences coming from that text type (table 10.2). This is actually somewhat surprising in the light of what was stated above. One would expect that the most basic of equivalents would be used less in texts requiring more variation to make them enjoyable reading. Looking at this from another perspective, of all of the fiction occurrences, *tappaa* covered ~71% while in non-fiction the percentage was only ~53% (table 10.3), a significantly smaller figure. The situation becomes

especially interesting when *tappaa* is compared with *surmata*, which was the second least common equivalent in fiction with ~38% (table 10.2). It could be that none of the other criteria that steer the choice toward *surmata* occur regularly in fiction, but this is not very likely, as those criteria include the figurative and passive uses of *kill*. On the other hand, there might be something in specific non-fiction texts types that makes the choice of *surmata* more likely. This will be discussed further in chapter 6.5.2 below. The following are examples of *tappaa* and *surmata* used in fiction, showing that these words are both used in surprisingly similar situations.

"You have **killed** my love," he muttered.
(WLD 7:155)
"Sinä olet **tappanut** minun rakkauteni", hän mutisi.

"Because each man **kills** the thing he loves," he said, "which was where we began."
(WEL 16:64)
"Koska rakkaansa **surmaa** joka mies", Carl May sanoi; "siitähän me aloitimme."

Of the remaining three, the ”other” category is the next one in this descending order of commonness in fiction. It is somewhat difficult to say anything about this category in this case, as it was almost evenly divided between fiction and non-fiction, with ~48% of it being in fiction (table 10.2). Perhaps noteworthy was that two out of the three cases of omitting the equivalent while keeping the sentence otherwise intact occurred in fiction texts, as shown in these examples:

Jälkeenpäin kuulin heidän lyöneen vetoja, kuka naisista kestäisi teloituksen tyynesti, mikä oli heidän järjestyksensä [**omitted**], mitkä viimeiset sanansa.
(LAN 14:36)
I heard afterwards that they had made bets on which of the women would calmly suffer the beheading, in what order they would be **killed**, and what their last words would be.

Jos ottaa kymmenenprosenttisen riskin [**omitted**], vääpelin laskelmien mukaan saa mennä seitsemän sekuntia.
(MER 4:346)
Seven seconds is to take a ten to one chance of getting **killed**:

In cases like these, it might be a stylistic choice to omit the exact word. The reader is expected to be able guess what is talked about, interpret it from the context. In many cases, non-fiction needs to be more precise to leave no room for this guesswork even if it results in more stilted texts. The third case of omission occurred in a sentence pair from David Attenborough's *Life of Birds*, which while non-fiction, seems to flow almost like a novel based on the data. This sentence pair is below.

> Often it is far too big for them to do so, for a peregrine may take a grouse and an eagle **kill** a hare.
> (ATT 4:166)
> Usein saalis olisi siihen liian suurikin, sillä muuttohaukka voi siepata teeren ja kotka **[omitted]** jäniksen.

6.5.2. More specific non-fiction types

First of all, it should noted that the figures and percentages here are based on only the part of the data that was categorized as non-fiction, not all of the data. This was done so that the non-fiction text types could be compared among each other more easily. Unfortunately, this leads to these figures not being comparable with those in any other tables in the corpus analysis chapter, at least not directly, as those are based on all of the data. Though because the word counts of the different text types cannot be counted, no frequencies can be figured out either, making the figures in this chapter guidelines at most in any case.

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
News	40	15	10	1	0	6	72
Art book	1	3	0	0	0	1	5
Grammar	2	0	0	1	0	0	3
History	1	2	0	0	1	1	5
Nature documentary	8	0	1	0	0	1	10
Scientific text	0	0	1	0	0	0	1
Other	2	1	0	0	1	2	6
Total	54	21	12	2	2	11	102

Table 10.4 – Non-fiction text types

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other
News	~74%	~71%	~83%	50%	0%	~55%
Art book	~ 2%	~14%	0%	0%	50%	~9%
Grammar	~ 4%	0%	0%	50%	0%	0%
History	~ 2%	~10%	0%	0%	50%	~9%
Nature documentary	~15%	0%	~8%	0%	0%	~9%
Scientific text	0%	0%	~8%	0%	0%	0%
Other	~ 4%	~ 5%	0%	0%	50%	~18%
Total	100%	100%	100%	100%	100%	100%

Table 10.5 – Non-fiction percentages according to the equivalent

	tappaa	*surmata*	*kuolla*	*henki*	*kaatua*	other	Total
News	~56%	~21%	~14%	~ 1%	0%	~8%	100%
Art book	20%	60%	0%	0%	0%	20%	100%
Grammar	~67%	0%	0%	~33%	0%	0%	100%
History	20%	40%	0%	0%	20%	20%	100%
Nature documentary	80%	0%	10%	0%	0%	10%	100%
Scientific text	0%	0%	100%	0%	0%	0%	100%
Other	~33%	~17%	0%	0%	~17%	~33%	100%

Table 10.6 – Non-fiction percentages according to the text type

The non-fiction category was sub-categorized mostly to see if news texts would show differences from the other text types and also to analyse the news texts a bit more on their own, as I find them an interesting to examine from the point of view of coarseness aspect of separating *kill* and *surmata* first mentioned in chapter 4.1.3. Unfortunately, examination of the first of these goals proved difficult, as no other text types besides news were nearly as common in the data as ”news”texts were, with 72 versus 10 (nature documentary) as the next most common type (table 10.4). Based on these results it might seem that news texts were disproportionally represented in TamBiC but this is more likely a case of *kill* being a word not likely to appear in text types such as cook books or grammar guides. On the other hand, someone being killed is almost always news-worthy and thus the word is likely to appear in this text type more often than in many others. Unfortunately, TamBiC does not provide a word count for the source texts, so it is impossible to find out the actual figures and see how much of news there are in the data compared to the other types.

To some extent, the prevalence of news texts could explain the relatively large number of occurrences of *surmata*. There were 15 of them (table 10.4), which is almost half of the total number of *surmata* occurrences (34, table 10.1). As was mentioned in chapter 4.1.3, according to NSS *surmata* is less coarse than *tappaa*. When there are news about killing, the objects of that killing are most often people. In fact, by cross-referencing between the results of this chapter and those of chapter 6.3.1, I found that almost none of the news texts which used *surmata* had a non-human object. This kind of situation requires some degree discretion, and therefore *surmata* is a logical choice of verb. However, this does not explain away the number of sentences where *tappaa* (with a human object) was used in news texts. It could be that in those instances discretion was not considered to be as necessary as in the *surmata* sentences. If this criterion is as important as it seems to be, it would make a thought-provoking problem for example for translation. Discretion is often difficult, and depends much on what a specific person considers to be a delicate subject. The

following sentences give examples of both *surmata* and *tappaa* used in news text in an unexpected way:

> On screen, according to fans, Arnold Schwarzenegger has *killed* 289 people, often with exceptionally large guns such as the Terminator's rocket-propelled grenade launcher.
> (ST 2003 8:6:3)
> Valkokankaalla Arnold Schwarzenegger on omien faniensa mukaan *surmannut* 289 ihmistä, usein erityisen suurilla aseilla kuten 'terminaattorin' kranaatteja laukaisevalla singolla.

> A Florida mother allegedly *killed* her 6-year-old son and fired a shotgun blast at another son whose life may have been saved by the Bible he was carrying.
> (REUTERS 2:2:3)
> Floridalaisen äidin oletetaan *tappaneen* kuusivuotiaan poikansa ja ampuneen haulikolla laukauksen kohti toista poikaansa, joka säilytti henkensä mahdollisesti kantamansa Raamatun ansiosta.

It is difficult to see why the first situation should need more discretion than the second, as the people killed in it are fictional, and the context shows that this killing most likely occurred in action movies, where discretion is not how death is handled. The second situation, a mother killing her children, on the other hand would be one requiring delicate handling.

As previously mentioned, besides news, the other non-fiction text types proved to be surprisingly uncommon. Consequently, a statistical analysis focusing on the figures and percentages would feel inappropriate as method, and because of this, the following discussion will mostly focus on bringing forth some particularly thought-provoking cases.

Another surprise might be some of the categories that appear on the tables. ”Nature documentary” and ”history” one might expect to find when examining *kill*, as the subject material often focuses on killing, but ”art book”, ”grammar” and ”scientific text” are more unlikely text types for this topic. These occurrences can be explained, however. Four out of the five ”art book” occurrences discuss Akseli Gallen-Kallela's *Kalevala* paintings, where once again, the subject material often turns to killing, such as in the following:

Nuori Joukahainen ei voi jatkaa elämäänsä, ellei hyvitä Aino-sisarensa kuolemaa, ja siksi hän yrittää **surmata** vanhan Väinämöisen, jota pitää tekoon syypäänä.
(G-K 34:12)
The young Joukahainen cannot live until he has revenged the death of his sister Aino, so he has to try and **kill** Väinämöinen whom he holds responsible.

The one remaining ”art book” occurrence is an interesting one in that it is rare in more than one way.

"Nykyään taiteessa on konstruktio, ja luulen, että juuri se käsite on **tappanut** taiteen".
(SCH 2:47:6)
"Nowadays there's all this talk about construction, and I believe concepts have **killed** art."

First is the text type itself, but the sentence pair is also an example of an inanimate object, as discussed in chapter 6.3.1, and finally, while it does use *tappaa*, the most common of the equivalents, it uses it in a context where *tappaa* is not as common as it could be. It is in fact the only occurrence of *tappaa* among the five ”art book” ones (table 10.4), and all in all I find the choice of *tappaa* here fascinating. Due to its commonness, it feels like the most basic of the equivalents, and thus not connected to discussions of art. The fact that the example sentence is in quotations marks brings forth the question whether this is in fact a spoken quote written down. Spoken language has text types very different from those of written language, and while it would be interesting to see how the equivalents other than *tappaa* are used in spoken language, it is unfortunately beyond the scope of this study, and would likely require a very different approach, as no corpus of Finnish-English spoken material exists at the time of writing this study.

The ”grammar” occurrences are actually example sentences used in the grammars, such as the following:

Myrkky **tappoi** koiran.
(K-M 1:41)
"The dog was **killed** by poison."

This on its own is perhaps one of the most basic sentence constructions in the entire data, which would explain the use *tappaa*. However, the same source text does use a more complicated construction as well, the following with *henki* in it:

> Hirvi **puski** sonnin **hengiltä** edustaa tapausta, missä alemman predikaation objekti (puski sonnia) on siirretty ylemmän predikaation yhdysverbin puski hengiltä objektiksi.
> (K-M 1:144)
> Hirvi puski sonnin hengiltä "The moose **killed** the bull by butting" represents a case where the object of the lower predication (puski sonnia) is moved to the position of the object of the higher predication (puski hengiltä).

The example sentences in grammars and other books directly discussing language are an interesting case as a text type in that they are chosen to exemplify the structures discussed, and thus might showcase structures not as likely to be found in more natural language, such as in the second of the two example sentences.

7. Discussion

This chapter will summarize the main findings of the previous corpus analysis chapter. My first research question was discovering what Finnish equivalents *kill* has and what English equivalents *tappaa* has. For *tappaa*, it was relatively straightforward: *kill* was the equivalent almost nine times out of ten. No other equivalents occurred in the TamBiC data enough times to be significant (3 occurrences at the most). Because of this, the rest of the research questions do not apply for *tappaa*, as a more in-depth analysis was carried out only for *kill.* For *kill*, the situation was a bit more complicated. Already based on dictionaries, which gave it as a synonym, *surmata* was examined alongside *tappaa*. Besides those two, *kuolla,* constructions using *henki* and *kaatua* occurred enough for a closer look. An interesting thing about these less common equivalents was the fact they had a causative relationship with *tappaa*, especially *kuolla*, but also some of the *henki* constructions and *kaatua*. A full list of the equivalents of *tappaa* in the data can be found in Appendix 1.

To answer the next question, "Which phenomena affect the choice of these equivalents?", the following part of this chapter will be organized by the equivalent used. Below each equivalent will be given those cases where that particular equivalent is particularly likely to occur, or, especially in the case of *tappaa* as the most common equivalent, those cases where the equivalent is less likely to occur. These will be given in the order the phenomena were discussed in chapter six, though not all of them will be mentioned for every equivalent, as in some cases nothing deviating from the baseline figures in table 2 or otherwise noteworthy was found.

Tappaa

* Of the inflected forms, *tappaa* showed clear preference for the base form and the present participle, while it was less common than the baseline with the 3rd person singular and past participle.

* Passives were one of those cases where *tappaa* was less common than its baseline, though with *get*-passives, the other equivalents were not as close to *tappaa* as was case with *be*-passives.

* Perhaps because it cannot combine with other phenomena, such as passive use, intransitive *kill* shows a very strong preference for *tappaa*.

* With both semantic subjects and objects, if the type (human, animal, inanimate) was unclear, it was with *tappaa* in all cases but one.

* *Tappaa* was rarer with figurative *kill* than with the literal, giving more weight to considering the other equivalents with this use.

* Quite surprisingly, *tappaa* took a larger share out of fiction occurrences than non-fiction ones when one would expect more complicated or rare equivalents in fiction.

Surmata

* *Surmata* showed a clear tendency to occur with the past participle of *kill*, and did not occur with

the present participle at all.

* *Surmata* and passives had an interesting relationship with the construction *saada surmansa* used more than the verb *surmata* itself. This is due to the restrictions of Finnish passive constructions. In general passives were more common with *surmata* than its baseline.

* While they were uncommon in the data, *surmata* occurred more with inanimate semantic object than its baseline was.

* *Surmata* did not show a clear preference for any particular type of semantic object.

* While figurative *kill* was overall very rare in the data, *surmata* did occur with it enough for the connection to be noteworthy.

* As surprising as the commonness of *tappaa* in fiction was the uncommonness of *surmata* in this text type.

* On the other hand, prevalence of *surmata* in news texts can be seen as a connection to one of the few pieces of information given about the relationship between it and *tappaa* in the dictionaries examined, that of *surmata* being less coarse than *tappaa*.

Kuolla

* *Kuolla* had an even stronger connection to the past participle than *surmata* did.

* *Kuolla* showed a very strong preference for *kill* used in the passive. This is connected to the causative relationship between *kuolla* and *tappaa* and the same restrictions of Finnish passives that lead to the use of *saada surmansa*.

* Most of the instances of *kuolla* occurred with human semantic objects.

* Exactly half of the instances of *kuolla* in the data had no semantic object.

* *Kuolla* was the equivalent with the largest percentage of non-fiction occurrences.

Henki

* *Henki* showed a slight preference for the past tense of *kill*, but the slight deviation from the baseline can be attributed to the small number of occurrences in this category.

* *Henki* showed same types of tendencies for semantic objects as *kuolla*, with most of them being human.

* As *henki* did occur with figurative *kill* at all, this could be taken as a sign of preference.

* Possibly because of the various constructions using *henki* being more complicated than the straightforward *tappaa*, *henki* showed preference for use in fiction over non-fiction.

Kaatua

* *Kaatua* only occurred with the past participle of *kill*, making it a very specific case when choosing an equivalent.

* *Kaatua* also only occurred with *kill* in the passive, meaning that all of the cases of the past participle mentioned above were caused by the use of the passive.

* All semantic objects of *kaatua* in the data were humans.

* *Kaatua* had only one occurrence where it had a semantic subject at all. This is once again explained by it being used exclusively with *kill* in passive.

* *Kaatua was* only used with literal *kill*.

* While *kaatua* did occur with both fiction and non-fiction, it showed a preference for the former.

Other

* 3rd person singular was for some for now indeterminable reason missing from the ”other” category.

* When the equivalent of *kill* was omitted from an equivalent sentence otherwise left in the text, it was most likely to be done in fiction texts.

My third research question was what kind of criteria are most important in deciding between the equivalents: syntactical (inflections, passive), semantic (semantic objects & subjects, semantic uses) or otherwise (text types). Based on the corpus study, no definite answer can be given to this question. All three affected the choice of equivalent for *kill* to some extent. The clearest concentration of choices other than *tappaa* seems to revolve around the use of passive, however. The more frequent use of passive meant that the other equivalents also appeared with the past perfect and without semantic subjects more often. The only equivalent for which a clear-cut use could be found was *kaatua*, which is a very marginal case as an equivalent of *kill*, and thus this result will not be very useful.

All in all, *kill* and *tappaa* might not have been the best choices of words for a study of this kind. In the previous studies I encountered, the different equivalents corresponded quite well to specific uses of the source word, as was predicted by McIntosh (see chapter 1.2). With *kill*, none of the phenomenona examined provided any concrete results, as *tappaa* was always the most common equivalent, and for the others, the percentages rarely reached amounts where anything concrete could be said. With *tappaa*, there simply was not enough variation for an actual study, though the prevalence of *kill* as an equivalent is a concrete result on its own. It might be that the amount of data was simply too small, and a larger corpus would be needed to examine these words. On the other hand, the inconclusiveness of the results can be seen as a result in itself. While what McIntosh concluded in 1960, that the choice of an equivalent can be predicted by grammatical features, works for some words, it does not work for all of them. *Kill* is an example of the latter kind of word, one where perhaps the choice is more of a stylistic one.

Concerning the methods used, Lado's model might seem old-fashioned, having been written originally in 1957. There is also the question of it having been meant specifically for language teaching, which was not the focus of this study. However, his three steps (see chapter 3.3) are still

useful today, as they are simple and easily understood and because of this, leave room for the modern researcher to add their own interpretation to the model. A more detailed model that relies on the way language was viewed sixty years ago would probably not have worked as well as Lado's.

The use of TamBiC was somewhat problematic. It was chosen as the corpus used by necessity of finding a corpus for the language pair of English-Finnish, and lacks many of the helpful features larger and more established corpora tend to have these days, for example words counts for the source texts and Key Word in Context aligning of the results. There is also the problem of the size of 2 million words, small by modern standards, and unlikely to be expanded as the compiler of the corpus has retired. However, the fact that it has an easily accessible Internet interface is definitely on the side of TamBiC, as is the fact that one can import the results of ones searches onto ones computer or print them. TamBiC works for what it states it was made for in its user manual (https://www12.uta.fi/tambic/user_manual.html): cross-language research of various types. The things examined just have be chosen from those that are relatively common in their languages to compensate for the small size.

8. Conclusion

In conclusion, the results of this study were not very conclusive. *Tappaa* had too little variation in its equivalents for a more in-depth study. *Kill* on the other hand had many Finnish equivalents, with *tappaa* as the most common one, but none of the others (besides the very rare *kaatua*) had a specific circumstances where they occurred exclusively. In most cases, they did not differ significantly from their baseline in the data, and even when they did, *tappaa* still had the largest figures in most of those cases. But the very existence of these other equivalents means that the equivalence relationship cannot simply be said to be *kill=tappaa*. Potentially significant were the cases when *kill* was in the passive, as there all the other equivalents besides *tappaa* showed increase in occurrences. Passives however only covered one fourth of the data examined (61 of the 241 instances of *kill*, see

table 6.1).

In general, the results of multilingual corpus research can be applied in various ways. Johansson (2007, chapter 15.5) suggests different types of natural language processing, lexicography, translator training and the old mainstay of contrastive linguistics, foreign-language teaching). But because of these inconclusive results, it is difficult to say whether they could be applied anywhere, at least not easily. In the early stages of this study, the application of the results to help machine translation was going to be the main approach, but this was left out because of time constraints and the difficulty in integrating a field with very different theoretical background into the study.

Every possible approach cannot of course be covered in a study of this scope. Potential avenues for widening the scope would for example be a diachronic study of the various equivalents, especially whether for example *surmata* is more common in older texts, as a common reaction I got when explaining my what my thesis was about was that *surmata* feels ”old-fashioned” somehow. Another possibility would be a survey of a kind, where people would have to choose between different equivalents in different situations, such as having the same sentence in active and passive or where they would have to rate the grammaticality of sentences with different equivalents, in the ideal case also explaining their choices. This would be useful for providing more data besides TamBiC, and bring a less literary view into the study. Also interesting would be combining these two approaches, to see how well the data TamBic provides matches the language use of current people, whether things have changed in the years between the texts there being translated and the current situation. As Arppe and Järvikivi (2007, 132) point out, combining methods not only helps prove the results of one method, but also brings out entirely new results that enrich our linguistic experience.

9. Works cited

9.1. Primary sources

BNCWeb (A web-based interface to the British National Corpus)
bncweb.uta.fi
Accessed 3.4.2012

COBUILD = *Collins COBUILD Advanced Learner's Dictionary*. 2006. Ed. John Sinclair et al. Glasgow: HarperCollins Publishers.

Hurme, Raija, Maritta Pesonen and Olli Syväoja, eds. 1990. *Englanti-Suomi Suursanakirja*. Helsinki: Werner Söderström Osakeyhtiö.

Hurme, Raija, Riitta-Leena Malin and Olli Syväoja, eds. 1984. *Uusi Suomi-Englanti Suursanakirja*. Helsinki: Werner Söderström Osakeyhtiö.

KTS = *Kielitoimiston sanakirja*. 2006. Ed. Eija-Riitta Grönroos. Helsinki: Kotimaisten kielten tutkimuskeskus.

LDEL = *Longman Dictionary of the English Language*. 1995. Ed. Sue Engineer and Heather Gray. London: Penguin Group.

NSS = *Nykysuomen sanakirja*. 1973. Ed. Matti Sadeniemi. Suomalainen Kirjallisuuden Seura.

OALD = *Oxford Advanced Learner's Dictionary of Current English*. 2005. Ed. Sally Wehmeier. Oxford: Oxford University Press.

SSKS = *Suuri suomen kielen sanakirja*. 1992. Ed. Timo Nurmi. Jyväskylä: Big Sur Oy & Gummerus Kustannus Oy.

TamBiC (*The Tampere Bilingual Corpus of Finnish and English*).
https://www12.uta.fi/tambic/JTambic.html
Accessed 18.10.2011

9.2. Secondary sources

Altenberg, Bengt and Sylviane Granger. 2002. “Recent trends in cross-linguistic lexical studies”. In *Lexis in Contrast. Corpus-based approaches*, eds. Bengt Altenberg and Sylviane Granger, 3-48. Philadelphia: John Benjamins Publishing Company

Arppe, Antti and Järvikivi, Juhani. 2007. ”Every method counts: Combining corpus-based and experimental evidence in the study of synonymy”. *Corpus Linguistics and Linguistic Theory* 3, 2: 131-159.

Baker, Mona. 2011. *In Other Words: A Coursebook on Translation*. Oxon: Routledge.

Baker, Paul. 2009. “Introduction”. In *Contemporary Corpus Linguistics,* ed. Paul Baker, 1-8. London: Continuum International Publishing.

Barlow, Michael. 2008. “Parallel texts and corpus-based contrastive analysis”. In *Current Trends in Contrastive Linguistics : Functional and Cognitive Perspectives*, eds. Gonzales, Maria de los Angeles Gomez, J. Lachlan Mackenzie and Elsa M. Gonzales Alvarez, 101-122. Amsterdam: John Benjamins Publishing Company.

Cruse, David Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Edmonds, Philip and Hirst, Graeme. 2002. ”Near-synonymy and lexical choice.” *Computational Linguistics* 28, 2: 105-144.

European Parliament Proceedings Parallel Corpus
http://www.statmt.org/europarl/
Accessed 7.4.2012

Facchinetti, Roberta, ed. 2007. *Corpus Linguistics 25 Years on*. Amsterdam: Rodopi B. V.

Franssila, Katariina. 2010. *Rakkaalla lapsella on monta nimeä – Englannin manage-verbin suomen- ja saksankieliset vastineet Euroopan komission teksteissä.* Tampere: Tampere University Press.

Granger, Sylviane, Jacques Lerot and Stephanie Petch-Tyson, eds. 2003. *Corpus-based Approaches to Contrastive Linguistics and Translation Studies.* Amsterdam: Rodopi B.V.

Granger, Sylviane. 2003. ”The corpus approach: a common way forward for contrastive linguistics and translation studies?”. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, eds. Sylviane Grange, Jacques Lerot and Stephanie Petch-Tyson, 17-31. Amsterdam: Rodopi B.V.

Gonzales, Maria de los Angeles Gomez, J. Lachlan Mackenzie and Elsa M. Gonzales Alvarez, eds. 2008. *Current Trends in Contrastive Linguistics: Functional and Cognitive Perspectives.* Amsterdam: John Benjamins Publishing Company.

Huddleston, Rodney. 1988. *English Grammar: an Outline*. Cambridge: Cambridge University Press.

Johansson, Stig. 2007. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins Publishing Company.

Kenny, Dorothy. 2009. ”Equivalence”. In *Routledge Encyclopedia of Translation Studies*, eds. Mona Baker and Gabriela Saldanha, 96–99. London: Routledge.

Krzeszowski, Tomasz P. 1990. *Contrasting Languages: The Scope of Contrastive Linguistics*. Berlin: Walter de Gruyter.

Lado, Robert. 1957. *Linguistics Across Cultures: applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Laviosa, Sara. 2002. *Corpus-based Translation Studies.* Amsterdam: Rodopi B.V.

Leech, Geoffrey and Steven Fligelstone. 1992. “Computers and Corpus Analysis”. In *Computers and Written Texts*, ed. Christopher S. Butler, 115-140. Oxford: Basil Blackwell Ltd.

Leech, Geoffrey and Jan Svartvik. 2002. *A Communicative Grammar of English*. Harlow: Pearson Education Limited.

Lewandowska-Tomaszczyk, Barbara and Dziwirek, Katarzyna. 2010. *Complex Emotions and Grammatical Mismatches: A Contrastive Corpus-based Study.* Berlin: Walter de Gruyter.

McIntosh, Angus. 1960. ”A four-letter word in *Lady Chatterly's Lover*”. In *Patterns of Languages: Papers in General Descriptive and Applied Linguistics*, Ed. Angus McIntosh and M. A. K. Halliday, 151-164. London: Longmans.

Meyer, Charles. 2002. *English Corpus Linguistics: An introduction.* Port Chester: Cambridge University Press.

Palmer, F. R. 1976. *Semantics: A New Outline*. Cambridge: Cambridge University Press.

Pertovaara, Heidi. 2012. *The verbs Work and Play and their Finnish Equivalents*. Tampere: Tampere University Press.

Salminen, Saara. 2008. *A Corpus-based study on English and Swedish Near-synonyms: The Case of Environment, Circumstances and Surroundings, and Miljö, Omständigheter and Omgivning*. Tampere: Tampere University Press.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Stubbs, Michael. 1996. *Text and Corpus Analysis: Computer-assisted studies of language and culture.* Blackwell.

Suoniemi, Paula. 2006. *Translation of the 3rd Person Singular Pronouns in English and Finnish*. Tampere: Tampere University Press.

Svartvik, Jan. 2007. ”Corpus linguistics 25+ years on”. In *Corpus Linguistics 25 Years on*, ed. Roberta Facchinetti, 11-25. Amsterdam: Rodopi B. V.

Toury, Gideon. 1995. *Descriptive Translation Studies – and Beyond*. Amsterdam: John Benjamins Publishing Company .

VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen and Irja Alho. 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Internet-version, accessed 8.4.2012 Available at: http://scripta.kotus.fi/visk

Appendices

Appendix 1. - Closer look at the equivalent categories in the corpus analysis chapter

Equivalent	Number of instances
Tappaa	
Tappaa	139
Tappaminen	10
Tapattaa	1
Tappo	2
Surmata	
Surmata	26
Saada surmansa/surma	8
Kuolla	
Kuollut	3
Kuolema	3
Kuolla	12
Henki	
Päästä hengestään	1
Ottaa/iskeä/puskea /saada hengiltä	5
Menettää henkensä	2
Kaatua	
Kaatua	8
Other	
Murhata	1
Itsemurha	2
Ampua	5
Tuhota	1
Joutua uhriksi	1
Siltä varalta, että	1
Jäädä alle	1
Teurastaa	1
Mennä	1
Kaato	1
Menehtyä	1
Pistää	1
Omission	3

Appendix 2. - Corpus source text list

Finnish Originals

AHT – Helena Ahti *Kotoinen pitopöytä: suomalaista silmänruokaa* (1981) (cooking book)
FNR – Anna-Maija Tanttu et al. *Taivaallista tarjottavaa* (1998) (cooking book)
G-K – Eija Kämäräinen *Akseli Gallen-Kallela: Katsoin outoja unia* (1994) (art book)
HML – Pirjo Henttonen *Hämeenlinna* (1992) (history)
HRV – Elina Hirvonen *Että hän muistaisi saman* (2005) (fiction)
HS – *Helsingin Sanomat* (2001, 2008, 2009) (news)
HVK – Paavo Haavikko *Lumeton aika* (1964) (fiction)
JNS – Matti Joensuu *Pahan pappi* (2006) (fiction)
JPL – Eeva Joenpelto *Neito kulkee vetten päällä* (1955) (fiction)
JPV – Matti Joenpolvi *Aitoa hiusta* (1969) & *Syntymätöntä porsasta* (1975) (fiction)
JUT – Eino Jutikkala *Suomen historia* (1966) (history)
KAL – *Suomen Luonnonsuojeluliiton kalenterit* (2000-2006) (nature documentary)
KAR – Fred Karlsson *Suomen kielioppi* (1982) (grammar & linguistics)
K-M – Eeva Kangasmaa-Minn *Verbien sisäisestä aspektista* (1978) (grammar & linguistics)
KNK – Anita Konkka *Hullun taivaassa* (1988) (fiction)
LAN – Leena Lander *Lankeaa pitkä varjo* (1986) (fiction)
LIN – Väinö Linna *Tuntematon sotilas* (1954) (fiction)
LKS – Rosa Liksom *Yhden yön pysäkki. Unohdettu vartti. Tyhjän tien paratiisit.* (1985-9) (fiction)
MER – Veijo Meri *Manillaköysi* (1957) (fiction)
MSK – Various articles on music (music articles)
PEK – Toivo Pekkanen *Lapsuuteni* (1953) (fiction)
PSL – Arto Paasilinna *Ulvova mylläri* (1981) (fiction)
PTL – Kalle Päätalo *Koillismaa* (1960) (fiction)
SCH – Salme Sarjas-Korte *Helen Schjerfbeck: tie synteesiin.* (1992) (art book)
SIL – F. E. Sillanpää *Nuorena nukkunut* (1931) (fiction)
SIM – Kirsi Simonsuuri *Paholaispoika* (1986) (fiction)
TAR – Lauri Poropudas *150 tarinaa (Suomen Postin historia)* (2006) (history)
TDE – *Scientific dissertation abstracts* (scientific text)
TRA – Articles from various Finnish newspapers and magazines (news)
TRI – Antti Tuuri *Insinöörin kertomus* (1980) & *Pohjanmaa* (1982) (fiction)
TUO – *Suomen käänteissanakirja (johdanto)* (1971) (grammar & linguistics)
WAL – Mika Walteri *Sinuhe egyptiläinen* (1945) (fiction)
YLE – *Suomen Yleisradion uutiset* (2003, 2004, 2007, 2008) (news)

English Originals

AMS – Kingsley Amis *The Old Devils* (1986) (fiction)
ATT – David Attenborough *The Life of Birds* (1998) (nature documentary)
CAL – Arianna Strassinopoulos *Maria Callas: the Woman Behind the Legend* (1980) (other)
COR – S. Pit Corder *Introducing Applied Linguistics* (1973) (grammar & linguistics)
FLK – William Faulkner *Light in August* (1932) (fiction)
FOW – John Fowles *The French Lieutenant's Woman* (1969) (fiction)
GLD – William Golding *Lord of the Flies* (1954) (fiction)
GRE – Graham Greene *The Power and the Glory* (1940) (fiction)
GUI – James Mackay *The Guinness Book of Stamps: Facts and Feats.* (1982) (other)

GW – *The Guardian Weekly* (2003) (news)
HEM – Ernest Hemingway *The Old Man and the Sea* (1952) (fiction)
HKI – Anne Roston Korkeakivi *Helsinki: A City Journal* (1998) (other)
HWK – Neil Hardwick *Hardwick's Sauce* (1988) (other)
INS – Various instruction manuals (other)
LAW – D. H. Lawrence *Sons and Lovers* (1913) (fiction)
LES – Doris Lessing *Memoirs of a Survivor* (1974) (fiction)
MSC – *Articles from Gramophone magazine* (1996-2003) (music articles)
OLV – Jamie Oliver *The Return of the Naked Chef* (2000) (cooking book)
ORW – George Orwell *Nineteen Eighty-Four* (1949) (fiction)
OSB – John Osborne *Look Back in Anger* (1957) (fiction)
POM – Elizabeth Pomeroy *The Cookery Year* (1973) (cooking book)
REN – Patrick Bade *Renoir* (1989) (art book)
REU – *Reuters* (2001) (news)
SCI – Extracts from various scientific texts (scientific text)
SHF – R. C. Sherriff *Journey's End* (1929) (fiction)
SMI – Zadie Smith *White Teeth* (2000) (fiction)
ST2 – © *The Sunday Times* (2002) (news)
ST3 – © *The Sunday Times* (2003) (news)
STH – Vikram Seth *A Suitable Boy* (1993) (fiction)
TRV – G. M. Trevelyan *History of England* (1945) (history)
VGH – William Feaver *Van Gogh* (1990) (art book)
WDR – Thornton Wilder *Our Town* (1938) (fiction)
WEL – Fay Weldon *The Cloning of Joanna May* (1989) (fiction)
WLD – Oscar Wilde *The Picture of Dorian Gray* (1891) (fiction)